

On the inference about the spectral distribution of high-dimensional covariance matrix based on high-frequency noisy observations

Ningning Xia^{*} and Xinghua Zheng[†]

*School of Statistics and Management, Shanghai
Key Laboratory of Financial Information Technology,
Shanghai University of Finance and Economics
777 Guo Ding Road, China, 200433
e-mail: xia.ningning@mail.shufe.edu.cn*

*Department of information systems,
business statistics and operations management
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
e-mail: xhzheng@ust.hk*

Abstract: In practice, observations are often contaminated by noise, making the resulting sample covariance matrix a signal-plus-noise sample covariance matrix. Aiming to make inferences about the spectral distribution of the population covariance matrix under such a situation, we establish an asymptotic relationship that describes how the limiting spectral distribution of (signal) sample covariance matrices depends on that of signal-plus-noise-type sample covariance matrices. As an application, we consider inferences about the spectral distribution of integrated covolatility (ICV) matrices of high-dimensional diffusion processes based on high-frequency data with microstructure noise. The (slightly modified) pre-averaging estimator is a signal-plus-noise sample covariance matrix, and the aforementioned result, together with a (generalized) connection between the spectral distribution of signal sample covariance matrices and that of the population covariance matrix, enables us to propose a two-step procedure to consistently estimate the spectral distribution of ICV for a class of diffusion processes. An alternative approach is further proposed, which possesses several desirable properties: it is more robust, it eliminates the effects of microstructure noise, and the asymptotic relationship that enables consistent estimation of the spectral distribution of ICV is the standard Marčenko-Pastur equation. The performance of the two approaches is examined via simulation studies under both synchronous and asynchronous observation settings.

AMS 2000 subject classifications: Primary, 62H12, secondary, 62G99, tertiary, 60F15.

^{*}Research partially supported by GRF 606811 and 16305315 of the HKSAR, NSFC 11501348, Shanghai Pujiang Program 15PJ1402300, IRTSHUFE and the state key program in the major research plan of NSFC 91546202.

[†]Research partially supported by DAG (HKUST) and GRF 606811 and 16305315 of the HKSAR.

Keywords and phrases: High-dimension, high-frequency, integrated covariance matrices, Marčenko-Pastur equation, microstructure noise.

1. Introduction

1.1. Motivation

Covariance structure is of fundamental importance in multivariate analysis and applications. While in the classical low-dimensional setting, a usually unknown covariance structure can be estimated by the sample covariance matrix, in the high-dimensional setting, it is now well understood that the sample covariance matrix is not a consistent estimator. Furthermore, in many applications the observations are contaminated. Below, we explain one such setting that motivates this work. Similar situations arise in many other settings, especially in signal processing (see, e.g., [Dozier and Silverstein \(2007a\)](#), [El Karoui \(2010b\)](#), and [Hachem et al. \(2012\)](#)).

Our motivating question arises in the context of estimating the so-called integrated covariance matrix of the high-dimensional diffusion process and has applications to the study of stock price processes. More specifically, suppose that we have p stocks whose (latent) log price processes are denoted by (X_t^j) for $j = 1, \dots, p$. Let $\mathbf{X}_t = (X_t^1, \dots, X_t^p)^T$, where T denotes the transpose. A widely used model for (\mathbf{X}_t) is

$$d\mathbf{X}_t = \boldsymbol{\mu}_t dt + \boldsymbol{\Theta}_t d\mathbf{W}_t, \quad t \in [0, 1], \quad (1.1)$$

where $(\boldsymbol{\mu}_t) = (\mu_t^1, \dots, \mu_t^p)^T$ is a p -dimensional drift process, $(\boldsymbol{\Theta}_t)$ is a $p \times p$ matrix for any t called the covolatility process, and (\mathbf{W}_t) is a p -dimensional standard Brownian motion. Both $(\boldsymbol{\mu}_t)$ and $(\boldsymbol{\Theta}_t)$ can be stochastic and depend on the Brownian motion (\mathbf{W}_t) . The interval $[0, 1]$ is the time period of interest, say, one trading day (=six and a half hours). The integrated covariance (ICV) matrix refers to

$$\text{ICV} := \int_0^1 \boldsymbol{\Theta}_t \boldsymbol{\Theta}_t^T dt.$$

The ICV matrix, in particular, its spectrum (i.e., its set of eigenvalues), plays an important role in financial applications such as factor analysis and risk management.

A classical estimator of the ICV matrix is the so-called realized covariance (RCV) matrix, which relies on the assumption that (\mathbf{X}_t) could be observed at high frequency. More specifically, suppose that (\mathbf{X}_t) could be observed at time points $t_i = i/n$ for $i = 0, 1, \dots, n$. Then, the RCV matrix is defined as

$$\text{RCV} = \sum_{i=1}^n \Delta \mathbf{X}_i (\Delta \mathbf{X}_i)^T,$$

where

$$\Delta \mathbf{X}_i = \begin{pmatrix} \Delta X_i^1 \\ \vdots \\ \Delta X_i^p \end{pmatrix} := \begin{pmatrix} X_{t_i}^1 - X_{t_{i-1}}^1 \\ \vdots \\ X_{t_i}^p - X_{t_{i-1}}^p \end{pmatrix}$$

stands for the vector of log returns over the period $[(i-1)/n, i/n]$. The consistency and central limit theorems for the RCV matrix under such a setting and when the dimension p is fixed are well known; see, for example, Andersen and Bollerslev (1998), Andersen et al. (2001), Barndorff-Nielsen and Shephard (2002), Jacod and Protter (1998), Mykland and Zhang (2006), among others.

To obtain a better understanding of the above setting, it is instructive to connect it with the usual multivariate analysis setting. In the simplest case, when $\boldsymbol{\mu}_t \equiv 0$ and $\boldsymbol{\Theta}_t \equiv \boldsymbol{\Theta}$, we have (1) $\text{ICV} = \boldsymbol{\Theta}\boldsymbol{\Theta}^T$; (2) the returns $\Delta \mathbf{X}_i \stackrel{\mathcal{D}}{=} \mathbf{Z}_i/\sqrt{n}$, where $\mathbf{Z}_i \sim_{i.i.d.} N(0, \text{ICV})$, and $\text{RCV} \stackrel{\mathcal{D}}{=} \frac{1}{n} \sum_{i=1}^n \mathbf{Z}_i \mathbf{Z}_i^T$. In other words, this simplest setting is equivalent to an i.i.d. observation setting in multivariate analysis, where the sample covariance matrix is used to estimate the population covariance matrix. In general, this situation is more complicated because both $(\boldsymbol{\mu}_t)$ and $(\boldsymbol{\Theta}_t)$ can be stochastic and dependent on the underlying Brownian motion.

The so-called market microstructure noise presents another challenge. In practice, the observed prices are always contaminated versions of the latent prices, the error being referred to as market microstructure noise. Such noise is induced by various frictions in the trading process such as the bid-ask spread and the discreteness of price. Despite its small size, market microstructure noise accumulates at high frequency and badly affects inferences about the latent price processes. Liu et al. (2015) compare various volatility estimators and point out that microstructure noise is not negligible when the sampling frequency is higher than one observation per five minutes.

The following additive model has been widely adopted in recent studies on volatility estimation:

$$\mathbf{Y}_{t_i} = \mathbf{X}_{t_i} + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, n, \quad (1.2)$$

where $\mathbf{Y}_{t_i} = (Y_{t_i}^1, \dots, Y_{t_i}^p)^T$ denotes the observations and $\boldsymbol{\varepsilon}_i = (\varepsilon_i^1, \dots, \varepsilon_i^p)^T$ denotes the noise, which is i.i.d. independent of (\mathbf{X}_t) with $E(\boldsymbol{\varepsilon}_i) = 0$ and certain covariance matrix $\boldsymbol{\Sigma}_\varepsilon$. Observe that under (1.2), the observed log-returns $\Delta \mathbf{Y}_{t_i} := \mathbf{Y}_{t_i} - \mathbf{Y}_{t_{i-1}}$ relate to the true log-returns $\Delta \mathbf{X}_{t_i}$ by the following equation:

$$\Delta \mathbf{Y}_{t_i} = \Delta \mathbf{X}_{t_i} + \Delta \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, n, \quad (1.3)$$

where, as usual, $\Delta \boldsymbol{\varepsilon}_i := \boldsymbol{\varepsilon}_i - \boldsymbol{\varepsilon}_{i-1}$. We are therefore in a noisy observation setting in which the observations are contaminated by additive noise. Such a setting forms the basis of the current work.

One striking feature in (1.3) that differs from most noisy observation settings is that as the observation frequency n goes to infinity, the signal, namely, the true log-return $\Delta \mathbf{X}_{t_i}$ becomes diminishingly small, while the noise $\Delta \boldsymbol{\varepsilon}_i$ remains

being of the same order of magnitude. Therefore, the signal-to-noise ratio goes to 0. A direct consequence is that even when the dimension is $p = 1$, the optimal rate for estimating ICV is only $n^{1/4}$ instead of the usual \sqrt{n} ; see [Gloter and Jacod \(2001\)](#). In other words, due to the *dominance of noise over signal*, the “effective sample size” is only $O(\sqrt{n})$ rather than n . This can be clearly seen from the preaveraging method that we will explain in [Section 2.1](#).

While the problem above constitutes our main motivation for considering a signal-plus-noise observation setting, our results are not restricted to this particular application. Our first main result, [Theorem 2.1](#), applies to a general setting where the signal and noise are of the same order of magnitude.

1.2. Summary of main results

Our main goal is to make inferences about the spectral distribution of the underlying population covariance matrix, in the setting above, the ICV matrix, based on the noisy observations (\mathbf{Y}_{t_i}) as in [\(1.2\)](#). We provide two approaches, which we summarize as follows.

Approach I requires two steps. We shall introduce an intermediate matrix, \mathcal{A}_m as defined in [\(2.9\)](#) below. Think of the ICV matrix as the underlying population covariance matrix. Because our observations are contaminated, the resulting sample covariance matrix is a signal-plus-noise sample covariance matrix. The intermediate matrix \mathcal{A}_m is a sample covariance matrix based on only signals. The two steps are then

- Step 1: derive the Stieltjes transform of the spectral distribution of \mathcal{A}_m based on the signal-plus-noise sample covariance matrix.
- Step 2: based on the derived Stieltjes transform of \mathcal{A}_m in Step 1, further consistently estimate the spectral distribution of ICV.

The two steps rely on two asymptotic results, [Theorems 2.1](#) and [2.2](#), respectively. Roughly speaking, [Theorem 2.1](#) enables us to make inferences about the signal sample covariance based on noisy observations, and [Theorem 2.2](#) allows us to go further back to the population covariance matrix.

Approach II is more direct. It makes use of some special properties in the setting that we described in [Section 1.1](#). The properties allow us to asymptotically eliminate the effect of noise, thus saving us from Step I above and enabling us to take only one step, which relies on [Theorem 2.3](#). We also see below that Approach II is more robust, particularly in that it allows for rather general dependence structures in the noise process, both cross-sectional and temporal, and even dependence between the noise and price process. The drawback is that Approach II heavily relies on the setting in [Section 1.1](#), while Approach I can be applied to wider situations involving noisy observations.

In the simulation studies, we explain in detail how to generalize the algorithm proposed by [El Karoui \(2008\)](#) to implement the estimation procedure in practice. We can see that Approaches I and II both yield satisfactory estimates of the spectral distribution of the targeting ICV matrix. Other algorithms, such as

those introduced in [Mestre \(2008\)](#), [Bai et al. \(2010\)](#) and [Ledoit and Wolf \(2015\)](#), can also be adapted to our setting.

The rest of the paper is organized as follows. Section 2 explains the two approaches and the underlying theories. Section 3 demonstrates how to implement the two approaches in practice. Section 4 concludes. The proofs are given in the supplementary article [Xia and Zheng \(2017\)](#).

Notation. For any $p \times p$ Hermitian matrix Σ with eigenvalues $\lambda_1, \dots, \lambda_p$, its empirical spectral distribution (ESD) is defined as

$$F^\Sigma(x) := \frac{1}{p} \#\{j : \lambda_j \leq x\}, \quad \text{for } x \in \mathbb{R}.$$

The limit of ESD as $p \rightarrow \infty$, if it exists, is referred to as the limiting spectral distribution, or LSD for short; see, for example, the book [Bai and Silverstein \(2010\)](#). For any real matrix \mathbf{A} , $\|\mathbf{A}\| = \sqrt{\lambda_{\max}(\mathbf{A}\mathbf{A}^T)}$ denotes its spectral norm, where λ_{\max} denotes the largest eigenvalue. For any nonnegative definite matrix \mathbf{B} , $\mathbf{B}^{1/2}$ denotes its square root matrix. For any $z \in \mathbb{C}$, write $\Re(z)$ and $\Im(z)$ as its real and imaginary parts, respectively, and \bar{z} as its complex conjugate. For any distribution F , $m_F(\cdot)$ denotes its Stieltjes transform, which is defined as

$$m_F(z) = \int \frac{1}{\lambda - z} dF(\lambda), \quad \text{for } z \in \mathbb{C}^+ := \{z \in \mathbb{C} : \Im(z) > 0\}.$$

In particular, the Stieltjes transform of F^Σ above, denoted by $m_\Sigma(\cdot) := m_{F^\Sigma}(\cdot)$, is given by

$$m_\Sigma(z) = \frac{\text{tr}((\Sigma - z\mathbf{I})^{-1})}{p}, \quad \text{for } z \in \mathbb{C}^+,$$

where \mathbf{I} is the identity matrix. Finally, for any vector \mathbf{x} , $|\mathbf{x}|$ stands for its Euclidean norm.

2. Main results

2.1. Preliminary: pre-averaging method

The pre-averaging (PAV) method is introduced in [Jacod et al. \(2009\)](#), [Podolskij and Vetter \(2009\)](#), and [Christensen et al. \(2010\)](#) to deal with microstructure noise. Other approaches include the two/multi-scales estimators ([Zhang et al., 2005](#); [Zhang, 2006, 2011](#)), realized kernel ([Barndorff-Nielsen et al., 2008](#); [Barndorff-Nielsen et al., 2011](#)), and quasi-maximum likelihood method ([Xiu, 2010](#); [Aït-Sahalia et al., 2010](#)). We use a slight variant of the PAV approach in this work. First, choose a window length k . Then, group the intervals $[(i-1)/n, i/n]$ for $i = 1, \dots, 2k \cdot \lfloor n/(2k) \rfloor$ into $m := \lfloor n/(2k) \rfloor$ pairs of non-overlapping

windows, each of width $(2k)/n$, where $\lfloor \cdot \rfloor$ represents rounding down to the nearest integer. Introduce the following notation for any process $\mathbf{V} = (\mathbf{V}_t)_{t \geq 0}$,

$$\Delta \mathbf{V}_i = \mathbf{V}_{i/n} - \mathbf{V}_{(i-1)/n}, \quad \bar{\mathbf{V}}_i = \frac{1}{k} \sum_{j=0}^{k-1} \mathbf{V}_{((i-1)k+j)/n}, \quad \text{and} \quad \Delta \bar{\mathbf{V}}_{2i} = \bar{\mathbf{V}}_{2i} - \bar{\mathbf{V}}_{2i-1}. \quad (2.1)$$

With such notation, the observed return based on the pre-averaged price becomes

$$\Delta \bar{\mathbf{Y}}_{2i} = \Delta \bar{\mathbf{X}}_{2i} + \Delta \bar{\boldsymbol{\varepsilon}}_{2i}. \quad (2.2)$$

One key observation is that if k is chosen to be of order \sqrt{n} (which is the order chosen in [Jacod et al. \(2009\)](#), [Podolskij and Vetter \(2009\)](#), and [Christensen et al. \(2010\)](#)), then, in (2.2), the “signal” $\Delta \bar{\mathbf{X}}_{2i}$ and “noise” $\Delta \bar{\boldsymbol{\varepsilon}}_{2i}$ can be shown to be of the same order of magnitude. Observe that with such a chosen window width, the resulting number of windows is only of order \sqrt{n} ; hence our statement earlier that the effective sample size is only $O(\sqrt{n})$ and consequently, even in the one-dimensional case, the optimal rate of convergence for estimating ICV is only $\sqrt{\sqrt{n}} = n^{1/4}$.

2.2. Approach I

2.2.1. Step 1: From signal-plus-noise back to signal

Our starting point is the PAV matrix, which is defined as a multiple of the sample covariance matrix of $\Delta \bar{\mathbf{Y}}_{2i}$, the returns based on the pre-averaged prices:

$$\text{PAV} := 3 \sum_{i=1}^m (\Delta \bar{\mathbf{Y}}_{2i}) (\Delta \bar{\mathbf{Y}}_{2i})^T. \quad (2.3)$$

(Coefficient 3 is inherited from [Jacod et al. \(2009\)](#) and comes from the convergence (2.9) below.) This is slightly different from the estimator in [Jacod et al. \(2009\)](#), particularly in that there is no bias correction term involved. This is because (1) in the high-dimensional setting, even with the bias correction, the PAV is still inconsistent, just as in high-dimensional multivariate analysis the sample covariance matrix is inconsistent; and (2) our version of the PAV facilitates further analysis, which leads us all the way back to the target ICV.

The matrix PAV can be viewed as the sample covariance matrix based on observations $\Delta \bar{\mathbf{X}}_{2i} + \Delta \bar{\boldsymbol{\varepsilon}}_{2i}$, which model the situation of the information vector $\Delta \bar{\mathbf{X}}_{2i}$ being contaminated by additive noise $\Delta \bar{\boldsymbol{\varepsilon}}_{2i}$. [Dozier and Silverstein \(2007a\)](#) consider such signal-plus-noise sample covariance matrices as

$$\mathbf{S}_n = \frac{1}{n} (\mathbf{A}_n + \sigma_n \mathbf{E}_n) (\mathbf{A}_n + \sigma_n \mathbf{E}_n)^T,$$

where \mathbf{A}_n indicates a matrix consisting of signals, while \mathbf{E}_n , independent of (\mathbf{A}_n) , consists of i.i.d. noise. Let $\mathcal{A}_n := \mathbf{A}_n \mathbf{A}_n^T / n$ be the signal sample covariance matrix. Under certain regularity conditions, the authors show that if $F^{\mathcal{A}_n}$

converges to a probability distribution $F^{\mathcal{A}}$, then so does $F^{\mathbf{S}_n}$. They further show that the LSD of \mathbf{S}_n is determined by $F^{\mathcal{A}}$ in that its Stieltjes transform $m = m(z)$ uniquely solves the following equation

$$m = \int \frac{dF^{\mathcal{A}}(t)}{\frac{t}{1 + \sigma^2 y m} - (1 + \sigma^2 y m)z + \sigma^2(1 - y)}, \quad \text{for all } z \in \mathbb{C}^+, \quad (2.4)$$

where σ^2, y are given in Assumptions (A.ii) and (A.iv) below.

Our goal in this article, as in many other applications, is to make inferences about signals based on noisy observations; in this case, to make inferences about \mathcal{A}_n based on \mathbf{S}_n . This motivates us to investigate the problem from a different angle than Dozier and Silverstein (2007a). Unlike (2.4), which states how the LSD of \mathbf{S}_n depends on that of \mathcal{A}_n , we show how the LSD of \mathcal{A}_n depends on that of \mathbf{S}_n ; see equation (2.5) below. We further explain how such a relation enables us to consistently estimate the ESD of \mathcal{A}_n based on \mathbf{S}_n .

The relation that we establish is essentially an inverse relation of (2.4). Inverting such relations is in general notoriously difficult. For example, the Marcenko-Pastur equation, which is similar to equation (2.4) and describes how the LSD of the sample covariance matrix depends on that of the population covariance matrix, is established long time ago in Marčenko and Pastur (1967), but it was after more than forty years that researchers realized how the (unobservable) ESD of the population covariance matrix can be recovered based on the (observable) ESD of the sample covariance matrix ((El Karoui, 2008; Mestre, 2008; Bai et al., 2010; Ledoit and Wolf, 2015) etc). In particular, Mestre (2008) derived an inverse formula for estimating individual population eigenvalues, under the assumption that the population covariance matrix admits only finitely many distinct eigenvalues with known multiplicity. Our first result, Theorem 2.1 below, gives an inverse relation of (2.4) that allows the derivation of the ESD of \mathcal{A}_n based on that of \mathbf{S}_n , under rather general assumptions.

We impose the following assumptions on the underlying matrices. Assumptions (A.i) and (A.iii) are from Dozier and Silverstein (2007a); in particular, (A.i) is about the convergence of the ESD of the signal sample covariance matrix. Assumption (A.ii) allows the variance of noise to depend on n as in the case of PAV. Assumption (A.iv) is standard in the studies of random matrices.

- (A.i) \mathbf{A}_n is $p \times n$, independent of ε_n , and with $\mathcal{A}_n = (1/n)\mathbf{A}_n\mathbf{A}_n^T$, $F^{\mathcal{A}_n} \xrightarrow{\mathcal{D}} F^{\mathcal{A}}$, where $F^{\mathcal{A}}$ is a probability distribution with the Stieltjes transform denoted by $m_{\mathcal{A}}(\cdot)$;
- (A.ii) $\sigma_n \geq 0$ with $\lim_{n \rightarrow \infty} \sigma_n = \sigma \in [0, \infty)$;
- (A.iii) $\mathbf{E}_n = (\epsilon_{ij})$ is $p \times n$ with the entries ϵ_{ij} being i.i.d. and centered with unit variance; and
- (A.iv) $n = n(p)$ with $y_n = p/n \rightarrow y > 0$ as $p \rightarrow \infty$.

We now present our first result about how the LSD of \mathcal{A}_n depends on that of \mathbf{S}_n .

Theorem 2.1. Suppose that Assumptions (A.i)-(A.iv) hold. Then, almost surely, the ESD of \mathbf{S}_n converges in distribution to a probability distribution F . More-

over, if F is supported by a finite interval $[a, b]$ with $a > 0$ and possibly has a point mass at 0, then $F^{\mathcal{A}}$ can be identified as follows. For all $z \in \mathbb{C}^+$ such that $m_{\mathcal{A}}(z) \in D_{\mathcal{A}}(y, \sigma^2) := \{\xi \in \mathbb{C} : z(1 - y\sigma^2\xi)^2 - \sigma^2(y-1)(1 - y\sigma^2\xi) \in \mathbb{C}^+\}$, $m_{\mathcal{A}}(z)$ uniquely solves the following equation

$$m_{\mathcal{A}}(z) = \int \frac{\tau}{\frac{\tau}{1 - y\sigma^2 m_{\mathcal{A}}(z)} - z(1 - y\sigma^2 m_{\mathcal{A}}(z)) + \sigma^2(y-1)} dF(\tau). \quad (2.5)$$

Remark 2.1. The restriction on $m_{\mathcal{A}}(z)$ to be in $D_{\mathcal{A}}$ is such that the integral on the right hand side of (2.5) is well-defined. Note that because $m_{\mathcal{A}}(z) \rightarrow 0$ and $zm_{\mathcal{A}}(z) \rightarrow -1$ as $\Im(z) \rightarrow \infty$, $m_{\mathcal{A}}(z)$ does belong to $D_{\mathcal{A}}$ for all z with $\Im(z)$ sufficiently large. Furthermore, by the uniqueness of analytic continuation, knowing the values of $m_{\mathcal{A}}(z)$ for z with $\Im(z)$ sufficiently large is sufficient to determine $m_{\mathcal{A}}(z)$ for all $z \in \mathbb{C}^+$.

Let us explain how Theorem 2.1 can be used to make inferences about signals based on noisy observations.

- (i) In practice, we observe noisy observations and can compute \mathbf{S}_n and hence its ESD. We can then replace F in (2.5) with $F^{\mathbf{S}_n}$ and solve for $m_{\mathcal{A}_n}(z)$. The empirical version of (2.5) can be solved numerically using, for example, the R package “rootSolve.” The uniqueness of the solution to equation (2.5) is theoretically justified using analytic tools. For its empirical version, we prove that if $\widehat{m_{\mathcal{A}_n}}(z)$ solves the empirical version, then it is close to the true $m_{\mathcal{A}_n}$ (see Appendix D). This property guarantees that even if the empirical version of (2.5) admits multiple solutions, they are all close to the true one. Consequently, because $m_{\mathcal{A}_n}(z)$ fully characterizes the ESD of \mathcal{A}_n , the estimated $m_{\mathcal{A}_n}(z)$ enables us to consistently estimate the ESD. In the simulation studies, we explain in detail how to implement this procedure in practice.
- (ii) More importantly, to further estimate the ESD of the population covariance matrix, in the next step to be developed, we need $m_{\mathcal{A}_n}(z)$. Theorem 2.1 provides such a necessary input. This is an important outcome of establishing the inverse relation (2.5).

In practice, if we are only interested in estimating the spectral distribution of the population covariance matrix, then, because in the second step we only need $m_{\mathcal{A}_n}(z)$, there is actually no need to estimate the ESD of \mathcal{A}_n . In the simulation studies, we still include this part but only for the purpose of illustrating the application of Theorem 2.1.

We now apply Theorem 2.1 to our PAV matrix. As mentioned in the summary in Section 1.2, in Step 1, we relate the PAV matrix to an intermediate matrix \mathcal{A}_m defined as follows:

$$\mathcal{A}_m := 3 \sum_{i=1}^m \Delta \bar{\mathbf{X}}_{2i} \cdot (\Delta \bar{\mathbf{X}}_{2i})^T. \quad (2.6)$$

It differs from the PAV matrix in that it does not involve the noise and can be regarded as a signal sample covariance matrix. The assumptions under our setting analogous to (A.i)-(A.iv) for Theorem 2.1 are then as follows.

- (B.i) the ESD of \mathcal{A}_m converges to a probability distribution $F^{\mathcal{A}}$ with the Stieltjes transform denoted by $m_{\mathcal{A}}(z)$;
- (B.ii) the noise $(\boldsymbol{\varepsilon}_i)_{1 \leq i \leq n}$ are independent of (\mathbf{X}_t) and are i.i.d. with zero mean and covariance matrix $\boldsymbol{\Sigma}_e = \sigma_e^2 \mathbf{I}$ for some $\sigma_e > 0$ and $\sigma_e \rightarrow 0$ as $p \rightarrow \infty$;
- (B.iii) $k = \lfloor \theta \sqrt{n} \rfloor$ for some $\theta \in (0, \infty)$, and $m = \lfloor \frac{n}{2k} \rfloor$ satisfies $\lim_{p \rightarrow \infty} p/m = y$.

We then have the following Corollary as a direct consequence of Theorem 2.1.

Corollary 2.1. Suppose that for all p , (\mathbf{X}_t) is a p -dimensional process satisfying (1.1). Suppose also that Assumptions (B.i)-(B.iii) hold. Then, almost surely, the ESD of PAV defined in (2.3) converges to a probability distribution F . Moreover, if F is supported by a finite interval $[a, b]$ with $a > 0$ and possibly has a point mass at 0, then $F^{\mathcal{A}}$ can be identified as follows. For all $z \in \mathbb{C}^+$ such that $m_{\mathcal{A}}(z) \in D_{\mathcal{A}}(y, 3\theta^{-2}\sigma_e^2)$, $m_{\mathcal{A}}(z)$ uniquely solves the following equation

$$m_{\mathcal{A}}(z) = \int \frac{\tau}{\frac{\tau}{1 - 3y\theta^{-2}\sigma_e^2 m_{\mathcal{A}}(z)} - z(1 - 3y\theta^{-2}\sigma_e^2 m_{\mathcal{A}}(z)) + 3\theta^{-2}\sigma_e^2(y - 1)} dF(\tau). \quad (2.7)$$

Remark 2.2. Although Corollary 2.1 is stated for the case when noise components have the same standard deviations, it can readily be applied to the case when the covariance matrix $\boldsymbol{\Sigma}_e$ is a general diagonal matrix, say, $\text{diag}(d_1^2, \dots, d_p^2)$. To see this, let $d_{max}^2 = \max(d_1^2, \dots, d_p^2)$. We can then artificially add additional $\tilde{\boldsymbol{\varepsilon}}_i$ to the original observations, where $\tilde{\boldsymbol{\varepsilon}}_i$ are independent of $\boldsymbol{\varepsilon}_i$ and are i.i.d. with zero mean and covariance matrix $\tilde{\boldsymbol{\Sigma}}_e = \text{diag}(d_{max}^2 - d_1^2, \dots, d_{max}^2 - d_p^2)$. The noise components in the modified observations then have the same standard deviation d_{max} , and Corollary 2.1 can be applied. Note that the variances, d_1^2, \dots, d_p^2 , can be consistently estimated; see, for example, Theorem A.1 in Zhang *et al.* (2005). A similar remark applies to Theorem 2.1.

2.2.2. Step 2: From signal to population

Step 1 enables us to infer the spectral distribution of \mathcal{A}_m based on PAV. However, just as in high-dimensional multivariate analysis where the sample covariance matrix is not consistent, neither is \mathcal{A}_m . This is why we need this second step, which enables us to go further back to the ICV matrix.

First, we introduce some structural assumptions on the latent process \mathbf{X} in order to go further. Note that the term $\Delta \bar{\mathbf{V}}_{2i}$ in (2.1) can be written in a more

clear form by using the triangular kernel:

$$\begin{aligned}
\Delta \bar{\mathbf{V}}_{2i} &= \frac{1}{k} \sum_{j=0}^{k-1} (\mathbf{V}_{((2i-1)k+j)/n} - \mathbf{V}_{((2i-2)k+j)/n}) \\
&= \frac{1}{k} \sum_{j=0}^{k-1} \sum_{\ell=1}^k \Delta \mathbf{V}_{(2i-2)k+j+\ell} \\
&= \sum_{|j|<k} \left(1 - \frac{|j|}{k}\right) \Delta \mathbf{V}_{(2i-1)k+j}.
\end{aligned} \tag{2.8}$$

Based on this, it can be shown that if the dimension p is fixed, then, as $n \rightarrow \infty$,

$$\sum_{i=1}^m \Delta \bar{\mathbf{X}}_{2i} \cdot (\Delta \bar{\mathbf{X}}_{2i})^T \xrightarrow{p} \frac{\text{ICV}}{3}, \text{ hence } \mathcal{A}_m \xrightarrow{p} \text{ICV}. \tag{2.9}$$

It is also easy to verify that

$$\Delta \bar{\boldsymbol{\varepsilon}}_{2i} \stackrel{\mathcal{D}}{=} \sqrt{\frac{2}{k}} \mathbf{e}_i,$$

where \mathbf{e}_i s are i.i.d. random vectors with mean zero and covariance matrix $\boldsymbol{\Sigma}_e$.

Corollary 2.1 in Step 1 allows us to consistently estimate the ESD of \mathcal{A}_m . In light of the convergence (2.9), it would have been sufficient for us to make inferences about the ICV if the convergence (2.9) also held in the high-dimensional case. Unfortunately, this is not the case, and a further step to go from \mathcal{A}_m to ICV is needed. Such an inference is generally impossible, as can be seen in the following. ICV is an integral $\int_0^1 \boldsymbol{\Theta}_t \boldsymbol{\Theta}_t^T dt$. In the simple situation where $\boldsymbol{\mu}_t \equiv 0$ and $\boldsymbol{\Theta}_t$ is deterministic, the building blocks in defining \mathcal{A}_m , $\Delta \mathbf{X}_i$, are multivariate normals with mean 0 and covariance matrices $\int_{(i-1)/n}^{i/n} \boldsymbol{\Theta}_t \boldsymbol{\Theta}_t^T dt$. The bottom line is all the n covariance matrices, $\int_{(i-1)/n}^{i/n} \boldsymbol{\Theta}_t \boldsymbol{\Theta}_t^T dt$ for $i = 1, \dots, n$, could be very different from the ICV! We can easily change the n covariance matrices $\int_{(i-1)/n}^{i/n} \boldsymbol{\Theta}_t \boldsymbol{\Theta}_t^T dt$ and hence the distributions of $\Delta \mathbf{X}_i$ without changing ICV. And as both the dimension p and observation frequency n go to infinity, there is too much freedom in the underlying distributions, which makes inferences about ICV impossible. Certain structural assumptions are necessary to turn the impossible possible. The simplest is to assume that $\boldsymbol{\Theta}_t \equiv \boldsymbol{\Theta}$, in which case $\Delta \mathbf{X}_i$ are i.i.d. The apparent drawback of this assumption is that it cannot capture stochastic volatility, which is a stylized feature in financial data. The following class of processes, introduced in Zheng and Li (2011), accommodates both stochastic volatility and the leverage effect while still making the inference about ICV possible (and the theory is already much more complicated than the i.i.d. observation setting).

Definition 2.1. Suppose that (\mathbf{X}_t) is a p -dimensional process satisfying (1.1). We say that (\mathbf{X}_t) belongs to Class \mathcal{C} if, almost surely, there exist $(\gamma_t) \in D([0, 1]; \mathbb{R})$

and $\mathbf{\Lambda}$ a $p \times p$ matrix satisfying $\text{tr}(\mathbf{\Lambda}\mathbf{\Lambda}^T) = p$ such that

$$\mathbf{\Theta}_t = \gamma_t \mathbf{\Lambda}, \quad (2.10)$$

where $D([0, 1]; \mathbb{R})$ stands for the space of càdlàg functions from $[0, 1]$ to \mathbb{R} .

Remark 2.3. The convention that $\text{tr}(\mathbf{\Lambda}\mathbf{\Lambda}^T) = p$ is made to resolve the non-identifiability built into the formulation (2.10), in which one can multiply (γ_t) and divide $\mathbf{\Lambda}$ by a same constant without modifying the process $(\mathbf{\Theta}_t)$. It is thus not a restriction.

Class \mathcal{C} incorporates some widely used models as special cases:

- The simplest case is when the drift $\boldsymbol{\mu}_t \equiv 0$ and $\gamma_t \equiv \gamma$, in which case the returns $\Delta \mathbf{X}_i$ are i.i.d. $N(0, \gamma^2/n \cdot \mathbf{\Lambda}\mathbf{\Lambda}^T)$.
- More generally, again, when the drift $\boldsymbol{\mu}_t \equiv 0$ while (γ_t) is independent of the underlying Brownian motion (\mathbf{W}_t) , the returns $\Delta \mathbf{X}_i$ follow mixed normal distributions.
 - Mixed normal distributions, or their asymptotic equivalent form in the high-dimensional setting, elliptic distributions (see Section 2 of [El Karoui \(2013\)](#) for the asymptotic equivalence), have been widely used in financial applications. [McNeil *et al.* \(2005\)](#) state that “elliptical distributions ... provided far superior models to the multivariate normal for daily and weekly US stock-return data” and that “multivariate return data for groups of returns of a similar type often look roughly elliptical.”
 - More recently, El Karoui, in a series of papers ([El Karoui, 2009, 2010a, 2013](#)), studied the Markowitz optimization problem under the setting that the returns follow mixed normal/elliptic distributions.
- Furthermore, Class \mathcal{C} allows the drift $(\boldsymbol{\mu}_t)$ to be non-zero and more importantly, the (γ_t) process to be stochastic and even dependent on the Brownian motion (\mathbf{W}_t) that drives the price process, thus featuring the so-called leverage effect in financial econometrics. The leverage effect is an important stylized fact of financial returns and has drawn a great deal of attention in recent years; see, for example, [Aït-Sahalia, Fan and Li \(2013\)](#) and [Wang and Mykland \(2014\)](#).

Observe that if (\mathbf{X}_t) belongs to Class \mathcal{C} , then the ICV matrix

$$\text{ICV} = \int_0^1 \gamma_t^2 dt \cdot \check{\boldsymbol{\Sigma}}, \quad \text{where } \check{\boldsymbol{\Sigma}} = \mathbf{\Lambda}\mathbf{\Lambda}^T. \quad (2.11)$$

Furthermore, if the drift process $\boldsymbol{\mu}_t \equiv 0$ and (γ_t) is independent of (\mathbf{W}_t) , then, conditional on (γ_t) and using (2.8), we have

$$\Delta \bar{\mathbf{X}}_{2i} \stackrel{\mathcal{D}}{=} \sqrt{w_i} \check{\boldsymbol{\Sigma}}^{1/2} \mathbf{Z}_i,$$

where $\mathbf{Z}_i = (Z_i^1, \dots, Z_i^p)^T$ consists of independent standard normals and

$$w_i = \sum_{|j| < k} \left(1 - \frac{|j|}{k}\right)^2 \int_{\frac{(2i-1)k+j-1}{n}}^{\frac{(2i-1)k+j}{n}} \gamma_t^2 dt. \quad (2.12)$$

It follows that

$$\mathcal{A}_m = 3 \sum_{i=1}^m \Delta \bar{\mathbf{X}}_{2i} \cdot (\Delta \bar{\mathbf{X}}_{2i})^T \stackrel{\mathcal{D}}{=} 3 \sum_{i=1}^m w_i \check{\Sigma}^{1/2} \mathbf{Z}_i \mathbf{Z}_i^T \check{\Sigma}^{1/2}.$$

We now explain how to make further inferences about the ICV matrix based on \mathcal{A}_m . Doing so relies on another asymptotic result which relates \mathcal{A}_m to ICV.

We impose the following assumptions on the underlying process. They are inherited from Proposition 5 of [Zheng and Li \(2011\)](#), and we refer the readers to that article for further background and explanations. Observe in particular that Assumption (C.v) allows the covolatility process to be dependent on the Brownian motion that drives the price processes. Such dependence allows us to capture the leverage effect. Assumptions (C.iv) and (C.vi) concern the spectral norm of the ICV matrix. We do not require the norm to be bounded, which allows, for example, spike eigenvalues.

Assumption C:

- (C.i) For all p , (\mathbf{X}_t) is a p -dimensional process in Class \mathcal{C} for some drift process $\boldsymbol{\mu}_t = (\mu_t^1, \dots, \mu_t^p)^T$ and covolatility process $(\boldsymbol{\Theta}_t) = (\gamma_t \boldsymbol{\Lambda})$;
- (C.ii) there exists $C_0 < \infty$ such that for all p and all $j = 1, \dots, p$, $|\mu_t^j| \leq C_0$ for all $t \in [0, 1)$ almost surely;
- (C.iii) as $p \rightarrow \infty$, the ESD of $\check{\Sigma} = \boldsymbol{\Lambda} \boldsymbol{\Lambda}^T$ converges to a probability distribution \check{H} ;
- (C.iv) there exist $C_1 < \infty$ and $\kappa < 1/6$ such that for all p , $\|\check{\Sigma}\| \leq C_1 p^\kappa$ almost surely;
- (C.v) there exists a sequence of index sets \mathcal{I}_p satisfying $\mathcal{I}_p \subset \{1, \dots, p\}$ and $\#\mathcal{I}_p = o(p)$ such that (γ_t) may depend on (\mathbf{W}_t) but only on $(W_t^j : j \in \mathcal{I}_p)$;
- (C.vi) there exists $C_2 < \infty$ such that for all p and for all $t \in [0, 1)$, $|\gamma_t| \leq C_2$ almost surely, and additionally, almost surely, (γ_t) converges uniformly to a nonzero process (γ_t^*) that is piecewise continuous with finitely many jumps.

We then have the following result connecting \mathcal{A}_m with ICV.

Theorem 2.2. Suppose that Assumptions (C.i)-(C.vi) and (B.iii) hold, then as $p \rightarrow \infty$,

- (i) the ESDs of ICV and \mathcal{A}_m converge to probability distributions H and $F^{\mathcal{A}}$ respectively, where

$$H(x) = \check{H}(x/\zeta) \quad \text{for all } x \geq 0 \text{ with } \zeta = \int_0^1 (\gamma_t^*)^2 dt; \quad (2.13)$$

(ii) $F^{\mathcal{A}}$ and H are related as follows:

$$m_{\mathcal{A}}(z) = -\frac{1}{z} \int \frac{\zeta}{\tau M(z) + \zeta} dH(\tau), \quad (2.14)$$

where $M(z)$ and another function $\tilde{m}(z)$ uniquely solve the following equations in $\mathbb{C}^+ \times \mathbb{C}^+$

$$\begin{cases} M(z) &= -\frac{1}{z} \int_0^1 \frac{(\gamma_s^*)^2}{1 + y\tilde{m}(z)(\gamma_s^*)^2} ds, \\ \tilde{m}(z) &= -\frac{1}{z} \int \frac{\tau}{\tau M(z) + \zeta} dH(\tau). \end{cases} \quad (2.15)$$

Equation (2.14) in Theorem 2.2 forms the basis for us to further estimate the ESD of ICV. It involves an unknown function $M(z)$, which can be solved as follows. First note that multiplying $\tilde{m}(z)$ and $M(z)$ on both sides of the first and second equations in (2.15), respectively, yields

$$\begin{cases} M(z) \cdot \tilde{m}(z) &= -\frac{1}{yz} + \frac{1}{yz} \int_0^1 \frac{1}{1 + y\tilde{m}(z)(\gamma_s^*)^2} ds, \\ M(z) \cdot \tilde{m}(z) &= -\frac{1}{z} \int \frac{\tau M(z)}{\tau M(z) + \zeta} dH(\tau) = -\frac{1}{z} - m_{\mathcal{A}}(z), \end{cases}$$

where the last step is due to (2.14). It follows that

$$-\frac{1}{z} - m_{\mathcal{A}}(z) = -\frac{1}{yz} + \frac{1}{yz} \int_0^1 \frac{1}{1 + y\tilde{m}(z)(\gamma_s^*)^2} ds, \quad (2.16)$$

and $\tilde{m}(z) = -(1/z + m_{\mathcal{A}}(z))/M(z)$. Substituting the last expression of $\tilde{m}(z)$ into equation (2.16) yields

$$\int_0^1 \frac{M(z)}{M(z) - (\gamma_s^*)^2 y(z^{-1} + m_{\mathcal{A}}(z))} ds = 1 - y - yz m_{\mathcal{A}}(z). \quad (2.17)$$

$M(z)$ is then obtained by plugging in the $m_{\mathcal{A}}(z)$ that we derived in Step 1 into (2.17) and solving for the solution that is unique in \mathbb{C}^+ by Theorem 1 in Zheng and Li (2011).

Having solved $M(z)$, we can then utilize equation (2.14) to estimate the ESD of ICV by generalizing the algorithms in El Karoui (2008), Mestre (2008), Bai *et al.* (2010) and Ledoit and Wolf (2015) etc. The resulting estimate can be shown to be consistent by using an argument similar to that for Theorem 2 of El Karoui (2008). The estimation procedure is explained in detail in the simulation studies.

2.3. Approach II

The second step in Approach I involves the process (γ_s^*) which is unknown in practice. Estimating this process inevitably introduces an additional source of error. Motivated by this consideration, we draw ideas from Zheng and Li

(2011) and develop an alternative approach that overcomes this difficulty. It is also worth mentioning that the alternative approach allows for rather general dependence structures in the noise process, both cross-sectional and temporal, and even dependence between the noise and price process. The temporal dependence between microstructure noise has been documented in recent studies; see, e.g., Hansen and Lunde (2006), Ubukata and Oya (2009) and Jacod et al. (2017).

We shall define a matrix that is an extension of the time-variation adjusted RCV matrix introduced in Zheng and Li (2011) to our noisy setting. To start, fix an $\alpha \in (1/2, 1)$ and $\theta \in (0, \infty)$, and let $k = \lfloor \theta n^\alpha \rfloor$ and $m = \lfloor n/(2k) \rfloor$. The time-variation adjusted PAV matrix is then defined as

$$\mathcal{B}_m := 3 \frac{\sum_{i=1}^m |\Delta \bar{\mathbf{Y}}_{2i}|^2}{m} \cdot \sum_{i=1}^m \frac{\Delta \bar{\mathbf{Y}}_{2i} (\Delta \bar{\mathbf{Y}}_{2i})^T}{|\Delta \bar{\mathbf{Y}}_{2i}|^2} = 3 \frac{\sum_{i=1}^m |\Delta \bar{\mathbf{Y}}_{2i}|^2}{p} \tilde{\Sigma}, \quad (2.18)$$

where

$$\tilde{\Sigma} := \frac{p}{m} \sum_{i=1}^m \frac{\Delta \bar{\mathbf{Y}}_{2i} (\Delta \bar{\mathbf{Y}}_{2i})^T}{|\Delta \bar{\mathbf{Y}}_{2i}|^2}. \quad (2.19)$$

Note that here, the window length k has a higher order than in Theorem 2.2. The reason is that, after pre-averaging, the underlying returns are $O_p(\sqrt{k}/n)$ and the noises are $O_p(\sqrt{1/k})$. In Theorem 2.2, we balance the orders of the two terms by choosing $k = O(\sqrt{n})$; here we take $k = O(n^\alpha)$ for some $\alpha > 1/2$, which enables us to asymptotically eliminate the effect of noise.

Next, recall the concept of ρ -mixing coefficients.

Definition 2.2. Suppose that $U = (U_k, k \in \mathbb{Z})$ is a stationary time series. For $-\infty \leq j \leq \ell \leq \infty$, let \mathcal{F}_j^ℓ be the σ -field generated by the random variables $(U_k : j \leq k \leq \ell)$. The ρ -mixing coefficients are defined as

$$\rho(r) = \sup_{f \in \mathcal{L}^2(\mathcal{F}_{-\infty}^0), g \in \mathcal{L}^2(\mathcal{F}_r^\infty)} |\text{Corr}(f, g)|, \quad \text{for } r \in \mathbb{N},$$

where, for any probability space Ω , $\mathcal{L}^2(\Omega)$ refers to the space of square-integrable, Ω -measurable random variables.

We now introduce a number of assumptions. Assumption (D.i) below says that we allow for rather general dependence structures in the noise process, both cross-sectional and temporal. We actually do not put any restrictions on the cross-sectional dependence, and even dependence between the noise and price process is allowed. Note also that Jacod et al. (2017) provides an approach to estimate the decay rate of the ρ -mixing coefficients. Assumption (D.ii) concerns the dependence between the covolatility process and the Brownian motion that drives the price processes. Assumption (D.iv) is about the boundedness of individual volatilities.

- (D.i) For all $j = 1, \dots, p$, the noise (ε_i^j) is stationary, has mean 0 and bounded 4ℓ th moments, and has ρ -mixing coefficients $\rho^j(r)$ satisfying $\max_{j=1, \dots, p} \rho^j(r) = O(r^{-\ell})$ for some integer $\ell \geq 2$;

- (D.ii) there exist $0 \leq \delta_1 < 1/2$ and a sequence of index sets \mathcal{I}_p satisfying $\mathcal{I}_p \subset \{1, \dots, p\}$ and $\#\mathcal{I}_p = O(p^{\delta_1})$ such that (γ_t) may depend on (\mathbf{W}_t) but only on $(W_t^j : j \in \mathcal{I}_p)$;
- (D.iii) there exists $C_1 < \infty$ such that for all p , $|\gamma_t| \in (1/C_1, C_1)$ for all $t \in [0, 1)$ almost surely;
- (D.iv) there exists $C_2 < \infty$ such that for all p and all j , the individual volatilities $\sigma_t = \sqrt{(\gamma_t)^2 \cdot \sum_{k=1}^p (\Lambda_{jk})^2} \in (1/C_2, C_2)$ for all $t \in [0, 1]$ almost surely;
- (D.v) there exist $C_3 < \infty$ and $0 \leq \delta_2 < 1/2$ such that for all p , $\|\text{ICV}\| \leq C_3 p^{\delta_2}$ almost surely;
- (D.vi) the δ_1 in (D.ii) and δ_2 in (D.v) satisfy that $\delta_1 + \delta_2 < 1/2$;
- (D.vii) $k = \lfloor \theta n^\alpha \rfloor$ for some $\theta \in (0, \infty)$ and $\alpha \in [(3 + \ell)/(2\ell + 2), 1)$, and $m = \lfloor \frac{n}{2k} \rfloor$ satisfy $\lim_{p \rightarrow \infty} p/m = y > 0$, where ℓ is the integer in (D.i).

Remark 2.4. Careful readers may have noted that Assumptions (B.iii) and (D.vii) are mathematically incompatible, as Assumption (B.iii) requires $p = O(\sqrt{n})$ while Assumption (D.vii) requires $p = O(n^{1-\alpha})$ for some $\alpha \in (1/2, 1)$. The two assumptions are, however, perfectly compatible in practice when we deal with finite samples. Take the choices of (p, n, k) in the simulation study in Section 3.1 below for example. There, we take $(p, n) = (100, 23400)$. When applying Corollary 2.1 and Theorem 2.2, we take $k = \lfloor 0.5\sqrt{n} \rfloor = 76$, which leads to $y = p/\lfloor n/2k \rfloor \approx 0.7$ in Assumption (B.iii); when applying Theorem 2.3 below, we take $k = \lfloor 1.5n^{0.6} \rfloor = 627$, which gives $y = p/\lfloor n/2k \rfloor \approx 5.6$ in Assumption (D.vii).

We have the following convergence result connecting \mathcal{B}_m with ICV.

Theorem 2.3. Suppose that Assumptions (C.i), (C.ii), (C.iii), (C.vi), and (D.i)-(D.vii) hold. Then, as $p \rightarrow \infty$, the ESDs of ICV and \mathcal{B}_m converge almost surely to probability distributions H and $F^{\mathcal{B}}$, respectively, where H satisfies (2.13) and $F^{\mathcal{B}}$ is determined by H in that its Stieltjes transform, denoted by $m_{\mathcal{B}}(z)$, satisfies the following (standard) Marčenko-Pastur equation

$$m_{\mathcal{B}}(z) = \int_{\tau \in \mathbb{R}} \frac{1}{\tau(1 - y(1 + zm_{\mathcal{B}}(z))) - z} dH(\tau), \quad \text{for } z \in \mathbb{C}^+. \quad (2.20)$$

Theorem 2.3 states that the LSDs of ICV and \mathcal{B}_m are related via the Marčenko-Pastur equation. Several algorithms have been developed to consistently recover H by inverting the Marčenko-Pastur equation; see, for example, El Karoui (2008); Mestre (2008); Bai *et al.* (2010); Ledoit and Wolf (2015) etc. We can therefore consistently estimate the ESD of ICV by using these existing algorithms.

2.4. A remark about asynchronicity

In multivariate high-frequency data analysis, in addition to microstructure noise, there is another challenge due to asynchronous trading. In practice, different

stocks are traded at different times; consequently, the tick-by-tick data are not observed synchronously. There are several existing methods for synchronizing data, such as the refresh times (Barndorff-Nielsen et al., 2011) and previous tick methods (Zhang, 2011). Asynchronicity is less of an issue than microstructure noise. For example, as pointed out in Zhang (2011), asynchronicity does not induce bias in the two-scales estimator, and even the asymptotic variance is the same as if there is no asynchronicity. While a rigorous treatment is beyond the scope of this article, we expect our methods to work for asynchronous data as well. The reason, roughly speaking, is as follows. Take the previous tick method for example. Here, we choose a (usually equally spaced) grid of time points $0 = t_0 < t_1 < \dots < t_n = 1$, and for each stock j , for each time point t_i , let τ_i^j be the latest transaction time before t_i . One then acts as if one observes $Y_{\tau_i^j}^j$ at time t_i for stock j . With the original additive model at time τ_i^j :

$$Y_{\tau_i^j}^j = X_{\tau_i^j}^j + \varepsilon_i^j,$$

we have at time t_i ,

$$Y_{t_i}^j := Y_{\tau_i^j}^j = X_{t_i}^j + \left((X_{\tau_i^j}^j - X_{t_i}^j) + \varepsilon_i^j \right). \quad (2.21)$$

In other words, the asynchronicity induces an additional error $(X_{\tau_i^j}^j - X_{t_i}^j)$. The error is, however, diminishingly small as the sampling frequency $n \rightarrow \infty$ because $X_{\tau_i^j}^j - X_{t_i}^j = O_p\left(\sqrt{t_i - \tau_i^j}\right) = o_p(1)$. In short, asynchronicity induces an additional error (and violates our model assumption); fortunately, the error is of negligible order compared with the microstructure noise (ε_i^j) . We therefore keep our focus on the model (1.3). In simulation studies, in addition to the synchronous observation setting, we consider an asynchronous setting where the observation times for different stocks are independent Poisson processes. We shall see that our methods still work well (see Section 3.2 for more details).

3. Simulation studies

In this section, we demonstrate how to estimate the ESD of ICV by using Approach I, which uses the PAV matrix, and Approach II, which uses the alternative matrix \mathcal{B}_m .

3.1. When observations are synchronous

We first consider a setting where observations are synchronous. To generate the underlying process \mathbf{X} , the process (γ_t) in Definition 2.1 is taken to be a stochastic U-shaped (γ_t) process as follows:

$$d\gamma_t = -\rho(\gamma_t - \phi_t) dt + \sigma d\widetilde{W}_t, \quad \text{for } t \in [0, 1],$$

where $\rho = 10$, $\sigma = 0.05$,

$$\phi_t = 2\sqrt{0.0009 + 0.0008 \cos(2\pi t)},$$

and $\widetilde{W}_t = \sum_{i=1}^p W_t^i / \sqrt{p}$ with W_t^i being the i th component of the Brownian motion (\mathbf{W}_t) that drives the price process. Observe that such a formulation makes (γ_t) dependent on *all* the component of the underlying Brownian motion; hence, Assumptions (C.v) and (D.ii) are both violated. However, we shall see that our methods still work well. A sample path of (γ_t) is given in Figure 7 in the supplementary article of [Xia and Zheng \(2017\)](#).

Next, the matrix $\check{\Sigma} = \Lambda \Lambda^T$ is taken to be $\mathbf{U} \mathbf{D} \mathbf{U}^T$, where \mathbf{U} is a random orthogonal matrix and \mathbf{D} is a diagonal matrix whose diagonal entries are drawn independently from the Beta(1, 3) distribution. Such generated $\check{\Sigma}$ does not necessarily have trace p , but as we pointed out in Remark 2.3, the assumption $\text{tr}(\check{\Sigma}) = p$ is a convention rather than a requirement. With such (γ_t) and $\check{\Sigma}$, the individual daily volatilities are around 3%, which is similar to what one observes in practice. The latent log price process (\mathbf{X}_t) follows

$$d\mathbf{X}_t = \gamma_t \Lambda d\mathbf{W}_t, \quad \text{where} \quad \Lambda = \check{\Sigma}^{1/2}. \quad (3.1)$$

Finally, the noise $(\varepsilon_i)_{1 \leq i \leq n}$ is taken to be i.i.d. $N(0, 0.0002 \mathbf{I})$.

In the studies below, the dimension, that is, the number of stocks p , is taken to be 100, and the observation frequency n is set to be 23,400, which corresponds to one observation per second on a regular trading day. Note again that because of the presence of noise, the “effective sample size” is only of order $\sqrt{n} \approx 153$, which is comparable to our chosen $p = 100$.

3.1.1. Estimation using Approach I: synchronous setting

We start with Approach I, which involves two steps.

In the first step, we replace F in equation (2.7) with the ESD of PAV and solve for $m_{\mathcal{A}}(z)$ using the R package “rootSolve.” The window length k in defining PAV is set to be $\lfloor 0.5\sqrt{n} \rfloor = 76$. As to the $m_{\mathcal{A}}(z)$ to be solved, we choose a set of z ’s whose real and imaginary parts are equally spaced in the intervals $[-20, 0]$ and $[1, 20]$, respectively. Denote these z ’s by $\{z_j\}_{j=1}^J$ and the estimated $m_{\mathcal{A}_m}(z_j)$ by $\widehat{m_{\mathcal{A}_m}}(z_j)$. We then need to estimate the ESD of \mathcal{A}_m based on $\{\widehat{m_{\mathcal{A}_m}}(z_j)\}_{j=1}^J$, which we do as follows.

Inspired by the nonparametric estimation method proposed in [El Karoui \(2008\)](#), we approximate $F^{\mathcal{A}_m}$ with a weighted sum of point masses

$$F^{\mathcal{A}_m} \approx \sum_{k=1}^K w_k \delta_{x_k}, \quad (3.2)$$

where $\{x_1 < x_2 < \dots < x_K\}$ is a grid of points to be specified and the w_k s are weights to be estimated. To choose the grid $\{x_k\}_{k=1}^K$, naturally, we would

like $[x_1, x_K]$ to cover the support of $F^{\mathcal{A}_m}$, which is unknown. To overcome this difficulty, note that the support of the ESD of PAV always covers that of \mathcal{A}_m ; hence, we can choose x_k s to be equally spaced between 0 and the largest eigenvalue of PAV, and we are guaranteed that $[x_1, x_K]$ covers the support of $F^{\mathcal{A}_m}$.

Next we discuss how to estimate the weights $\{w_k\}$ in (3.2). Observe that the discretization (3.2) gives an approximate Stieltjes transform of $F^{\mathcal{A}_m}$ as $\sum_{k=1}^K \frac{w_k}{x_k - z}$. Let

$$e'_j := \widehat{m_{\mathcal{A}_m}}(z_j) - \sum_{k=1}^K \frac{w_k}{x_k - z_j}, \quad j = 1, \dots, J$$

be the approximation errors. The weights $\{w_k\}_{k=1}^K$ are then estimated by minimizing the approximation errors:

$$\arg \min_{(w_1, \dots, w_K)} \max_{j=1, 2, \dots, J} \max\{|\Re(e'_j)|, |\Im(e'_j)|\} \quad \text{subject to} \quad \sum_{k=1}^K w_k = 1 \text{ and } w_k \geq 0. \quad (3.3)$$

Next, in Step 2, we estimate the ESD of ICV. By plugging in the $\{\widehat{m_{\mathcal{A}_m}}(z_j)\}_{j=1}^J$ obtained in the first step and solving equation (2.17), we obtain $\{\widehat{M}(z_j)\}_{j=1}^J$. The estimation of the ESD of ICV is then conducted similarly as above as follows. Discretize the ESD of ICV as

$$F^{\text{ICV}} \approx \sum_{k=1}^K c_k \delta_{x_k}, \quad (3.4)$$

where c_k s are again weights to be estimated. By equation (2.14), we expect that

$$e''_j := \widehat{m_{\mathcal{A}_m}}(z_j) + \frac{1}{z_j} \cdot \sum_{k=1}^K c_k \frac{\zeta}{x_k \widehat{M}(z_j) + \zeta}$$

to be small. The c_k s are then estimated by minimizing the approximation errors e''_j just as in (3.3).

Figures 1 and 2 below illustrate the estimation results. The left plot of Figure 1 shows three ESDs, those of ICV, \mathcal{A}_m , and PAV. The three curves are clearly different from each: the difference between PAV and \mathcal{A}_m is induced by noise, while that between \mathcal{A}_m and ICV is caused by high-dimensionality. Note that we only observe the ESD of PAV, whereas the ESDs of both ICV and \mathcal{A}_m are underlying. Our goal is to estimate the ESD of ICV. As we explained below Theorem 2.1, such a goal does not require estimating the ESD of \mathcal{A}_m . Here, we still estimate this ESD to illustrate the application of Corollary 2.1. The estimation of the ESD of \mathcal{A}_m is conducted in the first step, and the result is shown in the right plot of Figure 1. The second step estimates the ESD of ICV, with the result given in Figure 2.

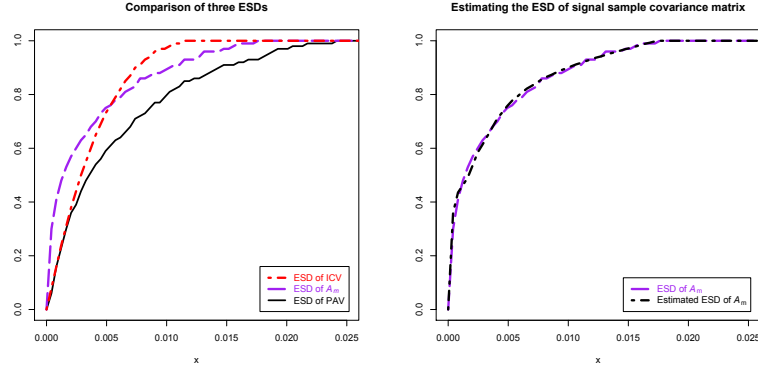


FIG 1. Approach I – Step 1: Estimation of the empirical spectral distribution of the signal sample covariance matrix \mathcal{A}_m based on synchronous noisy observations under model (3.1). The dimension $p = 100$, and the observation frequency $n = 23400$.

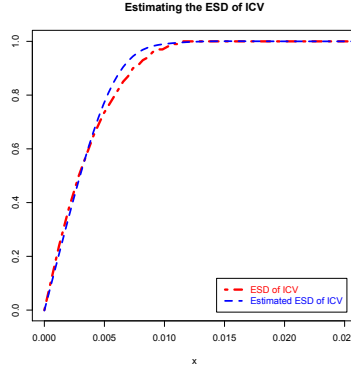


FIG 2. Approach I – Step 2: Estimation of the empirical spectral distribution of targeting ICV matrix.

Figures 1 and 2 show that the ESDs of both \mathcal{A}_m and ICV can be estimated quite well.

3.1.2. Estimation using Approach II: synchronous setting

We now apply Approach II to estimate the ESD of ICV. According to Theorem 2.3, asymptotically, the ESD of \mathcal{B}_m is related to that of ICV through the standard Marčenko-Pastur equation. This allows us to directly apply existing algorithms that are developed to invert the Marčenko-Pastur equation to estimate the ESD of ICV, and in the below we adopt the algorithm proposed in El Karoui (2008).

Specifically, set the window length k in defining \mathcal{B}_m to be $\lfloor 1.5n^{0.6} \rfloor = 627$. Discretize the ESD of ICV as (3.4). According to Theorem 2.3, the Stieltjes transform of the ESD of \mathcal{B}_m , denoted by $m_{\mathcal{B}_m}(z)$, should approximately satisfy equation (2.20) with H replaced with the ESD of ICV. In other words, we again expect the approximation errors

$$e_j''' := m_{\mathcal{B}_m}(z_j) - \sum_{k=1}^K \frac{c_k}{x_k(1 - y(1 + z_j m_{\mathcal{B}_m}(z_j))) - z_j}$$

to be small. Thus, again, we estimate the weights c_k 's by minimizing the approximation errors e_j''' as in (3.3).

The estimation results are given in Figure 3. Again, we see from the left plot that the ESD of \mathcal{B}_m clearly differs from the (latent unobserved) ESD of ICV, yet the right plot shows that we can estimate this latent distribution well.

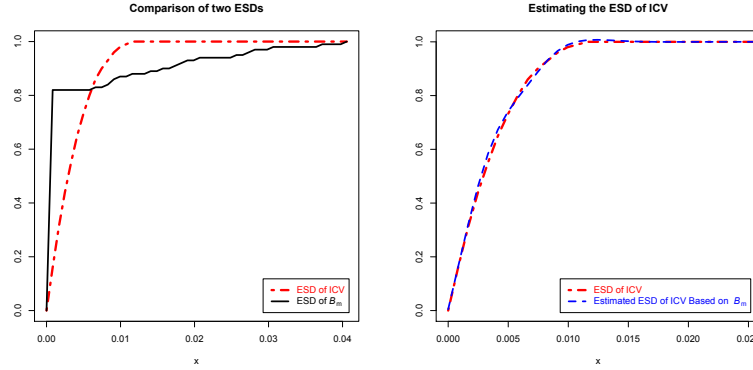


FIG 3. Approach II: Estimation of the empirical spectral distribution of the targeting ICV matrix based on synchronous noisy observations under model (3.1). The dimension $p = 100$ and the observation frequency $n = 23400$.

3.2. When observations are asynchronous

We now consider a setting where the observations are asynchronous. More specifically, for each stock $j = 1, \dots, p$, we simulate a Poisson process of rate 23,400 denoted by $\{\eta_i^j\}_{i=0,1,\dots}$. Because the Poisson processes $\{\eta_i^j\}_{i=0,1,\dots}$ are to be generated independently, almost surely, $\eta_{i_1}^{j_1} \neq \eta_{i_1}^{j_2}$ for all $1 \leq j_1 \neq j_2 \leq p$ and $0 < i_1, i_2$, namely, the observation times are all different for different stocks. Figure 8 in the supplementary article Xia and Zheng (2017) shows the observation times for three stocks generated in such a way during the first ten seconds. The observation times are highly irregularly spaced: there can be several seconds without a single observation, while there can also be several observations within a single second. Furthermore, because observation times for different stocks are generated independently, different stocks are observed in a rather

unsynchronized manner, making the estimation of covariances difficult. For this reason, a synchronization procedure needs to be carried out before we apply either Approach I or II.

Before we discuss how to synchronize data, we first continue with the simulation design. To generate the latent process (X_t) , because of the asynchronicity and high-dimensionality (we are dealing with $p = 100$ independent Poisson processes and, consequently, roughly $23,400 \times 100$ distinct observation times), there is a real technical difficulty in incorporating interactions among component processes in the data generating process. We adopt the following simplified setting to facilitate the simulation. Observe that the results in the previous subsection are achieved when the component processes have dependence, so we believe our methods would still work when there is asynchronicity *and* dependence. The simplified setting is as follows:

$$d\mathbf{X}_t = \phi_t \mathbf{D}^{1/2} d\mathbf{W}_t, \quad (3.5)$$

where both (ϕ_t) and \mathbf{D} are as in the previous subsection.

Our observations are

$$Y_{\eta_i^j}^j = X_{\eta_i^j}^j + \varepsilon_i^j,$$

where ε_i^j are i.i.d. $N(0, 0.0002)$.

Now we discuss how to synchronize data. We adopt the previous tick method explained in Section 2.4. More specifically, we choose an equally spaced grid $\{t_i\} \subseteq [0, 1]$, and for each t_i , for each $j = 1, \dots, p$, let

$$\tau_i^j = \max\{\eta_k^j : \eta_k^j \leq t_i\}.$$

We then proceed as if we observe $Y_{\tau_i^j}^j$ at time t_i . As we explained in Section 2.4, because $\tau_i^j \neq t_i$, such a synchronization procedure introduces an additional error $X_{\tau_i^j}^j - X_{t_i}^j$.

3.2.1. Estimation using Approach I: asynchronous setting

The additional error $X_{\tau_i^j}^j - X_{t_i}^j$ that the synchronization procedure induces depends on the latent process. For this reason, our independence assumption between the noise and the latent process (B.ii) is violated. To alleviate this problem, we synchronize less frequently so that the signals $X_{\tau_i^j}^j - X_{\tau_{i-1}^j}^j$ tend to be bigger and better approximate the true signals $X_{t_i}^j - X_{t_{i-1}}^j$. More specifically, we choose the equally spaced grid to be $\{t_i = 4i/23,400\}$ – in other words, we synchronize once every four seconds. Then, following the estimation procedure in Section 3.1.1, we have the following results.

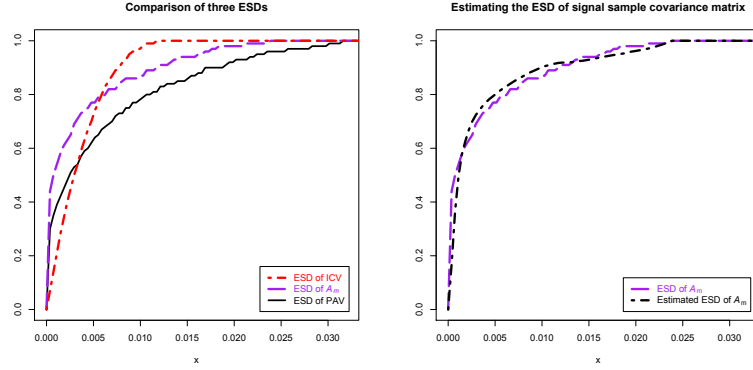


FIG 4. *Approach I – Step 1: Estimation of the empirical spectral distribution of the signal sample covariance matrix A_m based on asynchronous noisy observations under model (3.5). The dimension is $p = 100$. The synchronization frequency is 4 seconds, which leads to $n = 23400/4 = 5850$ observations.*

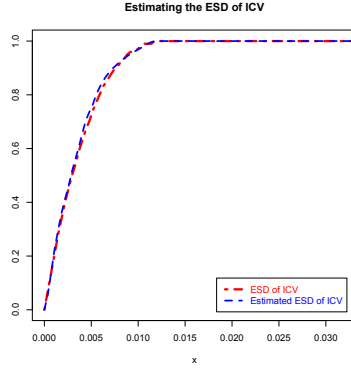


FIG 5. *Approach I – Step 2: Estimation of the empirical spectral distribution of the targeting ICV matrix.*

We see that in such a highly asynchronous noisy observation setting, Approach I still works quite well.

3.2.2. Estimation using Approach II: asynchronous setting

Approach II relies on Theorem 2.3, which allows for dependence in the noise process and between the noise and price process. For this reason, it is more robust than Approach I, and we can synchronize more frequently. In the estimation below we choose to synchronize once every second; that is, the time grid is taken to be $\{t_i = i/23400\}$. Then, following the estimation procedure in Section 3.1.2, we obtain the following results.

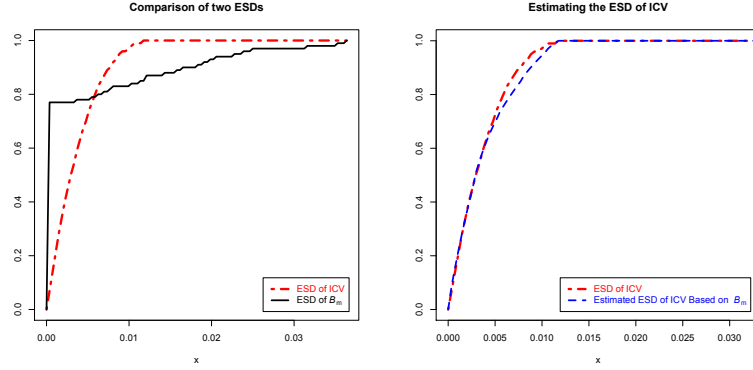


FIG 6. *Approach II: Estimation of the empirical spectral distribution of the targeting ICV matrix based on asynchronous noisy observations under model (3.5). The dimension is $p = 100$. The synchronization frequency is one second, which leads to $n = 23,400$ observations.*

Again, we see that in such an asynchronous noisy observation setting, Approach II works quite well.

3.3. Discussions about the two approaches

The two approaches have their own pros and cons.

- Approach I is more widely applicable to noisy observation situations. Moreover, in our particular application, because the window width in defining the PAV matrix is $O(\sqrt{n})$, which is of lower order than that for \mathcal{B}_m in Approach II, Approach I essentially has a larger “effective sample size.” This approach, however, is more sensitive to the model assumptions. In particular, in the asynchronous observation setting, because of the additional error introduced by asynchronicity, we may need to synchronize less frequently.
- Approach II is more direct because it only involves a one-step estimation procedure. It is also more robust because it allows for rather general dependence structures in the noise process, both cross-sectional and temporal, and even dependence between the noise and price process. For this reason, in the asynchronous setting in Section 3.2, we can use a higher synchronization frequency than for Approach I. A major drawback of Approach II is that it relies heavily on some special properties of the particular setting under study and hence may not be applicable to other noisy observation situations.

Finally, while in the estimation above we largely adapt the algorithms proposed by [El Karoui \(2008\)](#) to fit our setting, other algorithms such as those in [Mestre \(2008\)](#), [Bai et al. \(2010\)](#) and [Ledoit and Wolf \(2015\)](#) can also be adapted.

4. Conclusion

Motivated by the inference about the spectral distribution of the ICV matrix based on high-frequency noisy data,

- we establish an asymptotic relationship that describes how the spectral distribution of the signal sample covariance matrices depends on that of the sample covariance matrices constructed from noisy observations;
- using further a (generalized) connection between the spectral distribution of the signal sample covariance matrices and that of the population covariance matrix, we propose a two-step procedure that can consistently estimate the spectral distribution of ICV for a class of diffusion processes;
- we further develop an alternative approach that possesses several desirable properties: it is more robust, it eliminates the effects of microstructure noise, and the asymptotic relationship that enables the consistent estimation of the spectral distribution of ICV is the standard Marčenko-Pastur equation.
- numerical studies demonstrate that our proposed methods work well, under both synchronous and asynchronous observation settings.

Acknowledgements

We are very grateful to the editor, the associate editor, and the anonymous referees for their valuable comments and constructive suggestions, which led to a substantial improvement of this paper.

References

- Andersen, T. G., and Bollerslev, T. (1998). “Answering the skeptics: yes, standard volatility models do provide accurate forecasts.” *International Economic Review*, 39, 885–905.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2001). “The distribution of realized exchange rate volatility.” *Journal of the American Statistical Association*, 96, 42–55.
- Aït-Sahalia, Y., Fan, J. and Li, Y. (2010). “The leverage effect puzzle: Disentangling sources of bias at high frequency.” *Journal of Financial Economics*, 109, 224–249.
- Aït-Sahalia, Y., Fan, J. and Xiu, D. (2010). “High Frequency Covariance Estimates with Noisy and Asynchronous Financial Data.” *Journal of the American Statistical Association*, 105, 1504–1517.
- Bai, Z., Chen, J. and Yao, J. (2010). “On estimation of the population spectral distribution from a high-dimensional sample covariance matrix.” *Australian & New Zealand Journal of Statistics*, 52, 423–437.
- Bai, Z., and Silverstein, J. (2010). “Spectral analysis of large dimensional random matrices, 2nd edition.” *Springer, New York*.

- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde A. and Shephard N. (2008). "Designing realized kernels to measure ex-post variation of equity prices in the presence of noise." *Econometrica*, 76, 1481–1536.
- Barndorff-Nielsen, O. E., and Shephard N. (2002). "Econometric analysis of realized volatility and its use in estimating stochastic volatility models." *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 64, 253–280.
- Barndorff-Nielsen, O. E., Hansen, P.R., Lunde, A. and Shephard N. (2011). "Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading." *Journal of Econometrics*, 162, 149–169.
- Christensen, K., Kinnebrock, S. and Podolskij, M. (2010). "Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data." *Journal of Econometrics*, 159, 116–133.
- Dozier, R. and Silverstein, J. (2007a). "On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices." *Journal of Multivariate Analysis*, 98, 678–694.
- Dozier, R. and Silverstein, J. (2007b). "Analysis of the limiting spectral distribution of large dimensional information-plus-noise type matrices." *Journal of Multivariate Analysis*, 98, 1099–1122.
- El Karoui, N. (2008). "Spectrum estimation for large dimensional covariance matrices using random matrix theory." *Annals of Statistics*, 36, 2757–2790.
- El Karoui, N. (2009). "Concentration of measure and spectra of random matrices: applications to correlation matrices, elliptical distributions and beyond." *Annals of Applied Probability*, 19, 2362–2405.
- El Karoui, N. (2010a). "High-dimensionality effects in the Markowitz problem and other quadratic programs with linear constraints: risk underestimation." *Annals of Statistics*, 38, 3487–3566.
- El Karoui, N. (2010b). "On information plus noise kernel random matrices." *Annals of Statistics*, 38, 3191–3216.
- El Karoui, N. (2013). "On the realized risk of high-dimensional Markowitz portfolios." *SIAM J. Financial Math.*, 4, 737–783.
- Gloter, A. and Jacod, J. (2001). "Diffusions with measurement errors. ii - optimal estimators." *ESAIM*, 5, 243–260.
- Hachem, W., Loubaton, P., Mestre, X., Najim, J. and Vallet, P. (2012). "Large information plus noise random matrix models and consistent subspace estimation in large sensor networks." *Random Matrices: Theory Appl*, 1, 1150006.
- Hansen, P. R. and Lunde, A. (2006). "Realized variance and market microstructure noise." *Journal of Business and Economic Statistics*, 24, 127–161.
- Jacod, J., Li, Y., Mykland, P. A. Podolskij, M. and Vetter, M. (2009). "Microstructure noise in the continuous case: the pre-averaging approach." *Stochastic Process. Appl.*, 119, 2249–2276.
- Jacod, J., and Protter, P. (1998). "Asymptotic error distributions for the Euler method for stochastic differential equations." *Annals of Probability*, 26, 267–307.
- Jacod, J., Li, Y. and Zheng, X. (2017). "Statistical properties of microstructure noise." Available at SSRN: <http://ssrn.com/abstract=2212119>, to appear in

Econometrica.

- Ledoit, O., and Wolf, M. (2015). "Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions." *Journal of Multivariate Analysis*, 139, 360–384.
- Liu, L., Patton, A. and Sheppard, K. (2015). "Does anything beat 5-Minute RV? A comparison of realized measures across multiple asset classes." *Journal of Econometrics*, 187, 293–311.
- Marčenko, V. A. and Pastur, L. A. (1967). "Distribution of eigenvalues in certain sets of random matrices." *Mat. Sb. (N.S.)*, 72 (114), 507–536.
- McNeil, A. Frey, R. and Embrechts, P. (2005). "Quantitative risk management: concepts, techniques, and tools." *Princeton University Press*.
- Mestre, X. (2008). "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates." *IEEE Transactions on Information Theory*, 54, 5113–5129.
- Mykland, P. and Zhang, L. (2006). "ANOVA for diffusions and Ito processes." *Annals of Statistics*, 34, 1931–1963.
- Podolskij, M. and Vetter, M. (2009). "Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps." *Bernoulli*, 15, 634–658.
- Silverstein, J. and Bai, Z. (1995). "On the empirical distribution of eigenvalues of a class of large-dimensional random matrices." *Journal of Multivariate Analysis*, 54, 175–192.
- Ubukata, M. and K. Oya (2009). "Estimation and testing for dependence in market microstructure noise." *Journal of Financial Econometrics*, 7, 106–151.
- Wang, C. and Mykland, P. (2014). "The estimation of leverage effect with high-frequency data." *Journal of the American Statistical Association*, 109, 197–215.
- Xia, N. and Zheng, X. (2017). "Supplement to "On the inference about the spectral distribution of high-dimensional covariance matrix based on high-frequency noisy observations".
- Xiu, D. (2010). "Quasi-maximum likelihood estimation of volatility with high frequency data." *Journal of Econometrics*, 159, 235–250.
- Zhang, L. (2006). "Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach." *Bernoulli*, 12, 1019–1043.
- Zhang, L. (2011). "Estimating covariation: Epps effect, microstructure noise." *Journal of Econometrics*, 160, 33–47.
- Zhang, L., Mykland, P. and Aït-Sahalia, Y. (2005). "A tale of two time scales: determining integrated volatility with noisy high-frequency data." *Journal of the American Statistical Association*, 100, 1394–1411.
- Zheng, X. and Li, Y. (2011). "On the estimation of integrated covariance matrices of high dimensional diffusion processes." *Annals of Statistics*, 39, 3121–3151.

Ningning Xia: School of Statistics and Management, Shanghai, Key Laboratory of Financial Information Technology, Shanghai University of Finance and

Economics, 777 Guo Ding Road, China, 200433. xia.ningning@mail.shufe.edu.cn

Xinghua Zheng: Department of Information Systems Business Statistics and Operations Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. xhzheng@ust.hk

Supplementary Material

The proposition/lemma/equation etc. numbers below refer to the main article [Xia and Zheng \(2017\)](#).

Appendix A: Proof of Theorem 2.1

Theorem 2.1 is a consequence of the following proposition.

Proposition A.1. Under the assumptions of Theorem 2.1, there exists $K^* > 0$ such that almost surely, for all

$$z \in \mathbb{C}^* := \{z \in \mathbb{C}^+ : \Im(z) > K^*\},$$

we have

$$\lim_{p \rightarrow \infty} \left[\frac{1}{p} \text{tr} \left(\frac{1}{1 + \delta_n} \mathcal{A}_n - z \mathbf{I} \right)^{-1} - \frac{1}{p} \text{tr} \left(\mathbf{S}_n - (z - t_n \sigma_n^2) \mathbf{I} \right)^{-1} \right] = 0, \quad (\text{A.1})$$

where for all sufficiently large p , t_n is the unique solution to the equation

$$t_n = y_n - 1 + y_n(z - t_n \sigma_n^2) \frac{1}{p} \text{tr} (\mathbf{S}_n - (z - t_n \sigma_n^2) \mathbf{I})^{-1}, \quad (\text{A.2})$$

in the set

$$\mathcal{D} := \left\{ t \in \mathbb{C} : 0 \leq \Im(t) \leq \frac{\Im(z)}{2(\sigma + 1)^2} \right\}, \quad (\text{A.3})$$

and

$$\delta_n = y_n \sigma_n^2 \frac{1}{p} \text{tr} (\mathbf{S}_n - (z - t_n \sigma_n^2) \mathbf{I})^{-1}. \quad (\text{A.4})$$

In Section A.4 below, we show how Proposition A.1 leads to Theorem 2.1.

As to the proof of Proposition A.1, we shall use the following results from [Dozier and Silverstein \(2007a\)](#). By Theorem 1.1 therein, the sequence $\{F^{\mathbf{S}_n}\}$ converges weakly to a probability distribution F . Moreover, by using the same truncation and centralization technique as in [Dozier and Silverstein \(2007a\)](#), we may assume that

- (E.i) $|\epsilon_{11}| \leq a \log(n)$ for some $a > 2$,
- (E.ii) $E\epsilon_{11} = 0$, $E|\epsilon_{11}|^2 = 1$, and
- (E.iii) $\|(1/n)\mathbf{A}_n \mathbf{A}_n^T\| \leq \log(n)$.

In addition to equation (A.2), we shall also study its limiting equation

$$t = y - 1 + y(z - t\sigma^2)m(z - t\sigma^2), \quad (\text{A.5})$$

where $m(\cdot)$ is the Stieltjes transform of the probability distribution F .

It is shown in [Dozier and Silverstein \(2007b\)](#) that the distribution F admits a continuous density on $\mathbb{R} \setminus \{0\}$. Because we assume that F is supported by a finite interval $[a, b]$ with $a > 0$ and possibly has a point mass at 0, we conclude that F admits a bounded density f supported by $[a, b]$ and possibly a point mass at zero.

A.1. Properties of t_n and t

Lemma A.1. There exists $K_1 > 0$ such that for all $z \in \mathbb{C}_1 := \{z = u + iv : v > K_1\}$, for all sufficiently large n , equation (A.2) admits a unique solution in \mathcal{D} .

Proof. Rewrite equation (A.2) as

$$\begin{aligned} t_n + 1 &= y_n + y_n \int \frac{z - t_n \sigma_n^2}{x - z + t_n \sigma_n^2} dF^{S_n}(x) \\ &= y_n \int \frac{x}{x - z + t_n \sigma_n^2} dF^{S_n}(x). \end{aligned} \quad (\text{A.6})$$

First, under the assumptions of Theorem 2.1, by Theorem 1.1 in Bai and Silverstein (2012), if we let $[a_n, b_n]$ be an interval containing the support of F^{S_n} , then we may assume that for all large n , $b_n \leq \tilde{b} := b + 1$. Let $\tilde{\sigma} = \sigma + 1$, $\tilde{y} = y + 1$ and $K_1 = 2\tilde{\sigma}\sqrt{\tilde{y}\tilde{b}}$. Because $\sigma_n \rightarrow \sigma$ and $y_n \rightarrow y$, we have for all large n and for all $t = t_1 + it_2 \in \mathcal{D}$,

$$\sigma_n < \tilde{\sigma}, \quad y_n < \tilde{y}, \quad \text{and } v - t_2 \sigma_n^2 \geq v - t_2 \tilde{\sigma}^2 \geq v/2 > 0. \quad (\text{A.7})$$

Define

$$G(t) = y_n \int \frac{x}{x - z + t \sigma_n^2} dF^{S_n}(x) - 1, \quad \text{for all } t \in \mathcal{D}.$$

We apply the Banach fixed point theorem to show that for all sufficiently large n , there exists a unique point $t^* \in \mathcal{D}$ such that $G(t^*) = t^*$. The desired conclusion then follows.

Step (i): we prove that the mapping G is defined from \mathcal{D} to \mathcal{D} . From the definition of $G(t)$ and that of $t \in \mathcal{D}$, we have

$$\begin{aligned} \Im(G(t)) &= y_n \int_{a_n}^{b_n} \frac{x(v - t_2 \sigma_n^2)}{(x - u + t_1 \sigma_n^2)^2 + (v - t_2 \sigma_n^2)^2} dF^{S_n}(x) \\ &= \frac{y_n}{v - t_2 \sigma_n^2} \int_{a_n}^{b_n} \frac{x}{1 + \left(\frac{x - u + t_1 \sigma_n^2}{v - t_2 \sigma_n^2}\right)^2} dF^{S_n}(x), \end{aligned}$$

and hence for all sufficiently large n , by (A.7), we have

$$0 < \Im(G(t)) < \frac{\tilde{y}\tilde{b}}{v - t_2 \tilde{\sigma}^2} \leq \frac{v}{2\tilde{\sigma}^2},$$

where the last inequality follows from the fact that for any $z \in \mathbb{C}_1$,

$$\frac{\tilde{y}\tilde{b}}{v - t_2 \tilde{\sigma}^2} - \frac{v}{2\tilde{\sigma}^2} \leq \frac{2\tilde{y}\tilde{b}}{v} - \frac{v}{2\tilde{\sigma}^2} = \frac{4\tilde{\sigma}^2 \tilde{y}\tilde{b} - v^2}{2\tilde{\sigma}^2 v} \leq 0.$$

Step (ii): we shall show that $G : \mathcal{D} \rightarrow \mathcal{D}$ is a contraction mapping. In fact, for any two points $t, t' \in \mathcal{D}$,

$$\begin{aligned} G(t) - G(t') &= y_n \int_{a_n}^{b_n} \left(\frac{x}{x - z + t\sigma_n^2} - \frac{x}{x - z + t'\sigma_n^2} \right) dF^{S_n}(x) \\ &= (t - t') y_n \sigma_n^2 \int_{a_n}^{b_n} \frac{-x}{(x - z + t\sigma_n^2)(x - z + t'\sigma_n^2)} dF^{S_n}(x) \\ &:= (t - t') q(t, t'). \end{aligned}$$

By the Cauchy-Schwartz inequality and (A.7) we get that for all sufficiently large n , for all $t, t' \in \mathcal{D}$,

$$\begin{aligned} |q(t, t')| &\leq \left(\int_{a_n}^{b_n} \frac{\sigma_n^2 y_n x}{|x - z + t\sigma_n^2|^2} dF^{S_n}(x) \right)^{1/2} \left(\int_{a_n}^{b_n} \frac{\sigma_n^2 y_n x}{|x - z + t'\sigma_n^2|^2} dF^{S_n}(x) \right)^{1/2} \\ &\leq \left(\frac{\sigma_n^2 y_n b_n}{(v - \Im(t)\sigma_n^2)^2} \right)^{1/2} \left(\frac{\sigma_n^2 y_n b_n}{(v - \Im(t')\sigma_n^2)^2} \right)^{1/2} \\ &< \left(\frac{\tilde{\sigma}^2 \tilde{y} \tilde{b}}{(v - \Im(t)\tilde{\sigma}^2)^2} \right)^{1/2} \left(\frac{\tilde{\sigma}^2 \tilde{y} \tilde{b}}{(v - \Im(t')\tilde{\sigma}^2)^2} \right)^{1/2} \\ &\leq \left(\frac{\tilde{\sigma}^2 \tilde{y} \tilde{b}}{v^2/4} \right)^{1/2} \left(\frac{\tilde{\sigma}^2 \tilde{y} \tilde{b}}{v^2/4} \right)^{1/2}, \end{aligned}$$

which is strictly smaller than 1 when $z \in \mathbb{C}_1$. Therefore, the mapping G is contractive in \mathcal{D} , and the Banach fixed point theorem guarantees the existence of a unique solution to equation (A.2). \square

Lemma A.2. Suppose that t solves equation (A.5) for $z \in \mathbb{C}^+$. Write $t = t_1 + it_2$ and $z = u + iv$. Then, $0 < t_2 < v/\sigma^2$; moreover, as $v \rightarrow \infty$, uniformly in u , one has $t_2 \rightarrow 0$ and $t_1 \rightarrow -1$.

Proof. Taking imaginary parts on both sides of equation (A.5) yields

$$t_2 = y \int_a^b \frac{x(v - t_2\sigma^2)}{|x - z + t\sigma^2|^2} dF(x). \quad (\text{A.8})$$

It is then straightforward to verify that $t_2 > 0$ and $v - t_2\sigma^2 > 0$. Furthermore, since

$$\begin{aligned} t_2 &= \frac{y}{v - t_2\sigma^2} \int_a^b \frac{x}{1 + \left(\frac{x - u + t_1\sigma^2}{v - t_2\sigma^2} \right)^2} dF(x) \\ &\leq \frac{yb}{v - t_2\sigma^2}, \end{aligned} \quad (\text{A.9})$$

when $v \geq 2\sigma\sqrt{yb}$, we have

$$\text{either } t_2 \geq \frac{v + \sqrt{v^2 - 4\sigma^2 yb}}{2\sigma^2} \quad \text{or} \quad t_2 \leq \frac{v - \sqrt{v^2 - 4\sigma^2 yb}}{2\sigma^2}. \quad (\text{A.10})$$

Denote $w = u - t_1\sigma^2$ and $\theta = v - t_2\sigma^2$. By (A.9), if F admits a bounded density f and possibly a point mass at 0, then

$$\begin{aligned} t_2 &= \frac{y}{\theta} \int_a^b \frac{x}{1 + \left(\frac{x-w}{\theta}\right)^2} f(x) dx \\ &= y \int_{\frac{a-w}{\theta}}^{\frac{b-w}{\theta}} \frac{w + \theta l}{1 + l^2} f(w + \theta l) dl. \end{aligned}$$

Because $f(\cdot)$ is bounded and $x = w + \theta l \in (a, b)$ when $l \in (\frac{a-w}{\theta}, \frac{b-w}{\theta})$, there exists a constant C such that

$$t_2 \leq C \int_{\frac{a-w}{\theta}}^{\frac{b-w}{\theta}} \frac{1}{1 + l^2} dl \leq C \int_{-\infty}^{+\infty} \frac{dl}{1 + l^2} = C \pi.$$

This, combined with (A.10), implies that

$$t_2 \leq \frac{v - \sqrt{v^2 - 4\sigma^2 yb}}{2\sigma^2}, \quad \text{for all } v \text{ large enough.} \quad (\text{A.11})$$

In particular, uniformly in u ,

$$t_2 \rightarrow 0 \text{ and } v - t_2\sigma^2 \rightarrow \infty, \quad \text{as } v \rightarrow \infty. \quad (\text{A.12})$$

Moreover, from (A.5) we get

$$t + 1 = y + y \int \frac{z - t\sigma^2}{x - z + t\sigma^2} dF(x) = y \int \frac{x}{x - z + t\sigma^2} dF(x).$$

Thus as $v \rightarrow \infty$,

$$|t_1 + 1| \leq |t + 1| \leq y \int_a^b \frac{x}{|\Im(x - z + t\sigma^2)|} dF(x) \leq \frac{C}{v - t_2\sigma^2} \rightarrow 0,$$

also uniformly in u . □

Lemma A.3. There exists $K_2 \geq K_1$ such that for any $z \in \mathbb{C}_2 := \{z = u + iv : v > K_2\}$, equation (A.5) admits a unique solution.

Proof. First, by the same proof as for Lemma A.1, one can show that for all $z = u + iv$ with $v \geq K_1$, equation (A.5) admits a unique solution in \mathcal{D} . Moreover, by Lemma A.2, if $t = t_1 + it_2$ solves (A.5), then $t_2 > 0$; furthermore, we can

find a constant K_2 such that if t solves (A.5) for z with $v(= \Im(z)) \geq K_2$, then we must have $t_2 \leq v/(2\sigma^2)$. The latter two properties imply that for all z with $v \geq K_2$, the solution to (A.5) must lie in \mathcal{D} . Redefining $K_2 = \max(K_1, K_2)$ if necessary, we see that for all $z \in \mathbb{C}_2$, (A.5) admits a unique solution. \square

Lemma A.4. There exists $K_3 \geq K_2$ such that the solution $t = t(z)$ to (A.5) is analytic on $\mathbb{C}_3 := \{z = u + iv : v > K_3\}$.

Proof. Define a function G as

$$G(z, t) = t - (y - 1) - y(z - t\sigma^2)m(z - t\sigma^2), \quad (z, t) \in \mathbb{C}^+ \times \mathbb{C}^+ \text{ with } \Im(z - t\sigma^2) > 0.$$

That $t(z)$ solves (A.5) is equivalent to $G(z, t(z)) = 0$. Write $z = u + iv$ and $t = t_1 + it_2$. By taking the partial derivative with respect to t we get

$$\frac{\partial G}{\partial t} = 1 + y\sigma^2 \int \frac{x}{(x - (z - t\sigma^2))^2} dF(x).$$

Note that

$$\left| \int \frac{x}{(x - (z - t\sigma^2))^2} dF(x) \right| \leq \frac{b}{(v - t_2\sigma^2)^2},$$

which, by (A.12), goes to zero as $v \rightarrow \infty$. Thus there exists a constant $K_3 > 0$ such that for all $z \in \mathbb{C}_3$, $\partial G / \partial t(z, t(z)) \neq 0$. It follows from the implicit function theorem and Lemma A.3 that $t = t(z)$ is analytic on \mathbb{C}_3 . \square

Lemma A.5. Suppose that t_n solves equation (A.2) for $z \in \mathbb{C}_2$; then, $\Im(t_n) > 0$ and $\Im(z - t_n\sigma_n^2) > 0$. moreover, if t_n is the unique solution in the set \mathcal{D} , then, with probability one, as $n \rightarrow \infty$, t_n converges to a nonrandom complex number t that uniquely solves equation (A.5).

Proof. Write $z = u + iv$ and $t_n = t_{n1} + it_{n2}$. Similar to the proof of Lemma A.2, taking imaginary parts on both sides of equation (A.2), one can easily show that $t_{n2} > 0$ and $v - t_{n2}\sigma_n^2 > 0$.

Next we show that $\{t_n\}$ is tight; in other words, for any $\varepsilon > 0$, there exists $C > 0$, such that for all sufficiently large n , $P(|t_n| > C) < \varepsilon$. Because $0 < t_{n2} < v/\sigma_n^2$, it suffices to show that $\{|t_{n1}|\}$ is tight.

Let $\underline{\mathbf{S}}_n = \frac{1}{n}(\mathbf{A}_n + \sigma_n \mathbf{E}_n)^T (\mathbf{A}_n + \sigma_n \mathbf{E}_n)$, and let $\underline{m}_n(z)$ be the Stieltjes transform of the ESD $F^{\underline{\mathbf{S}}_n}$. The spectra of \mathbf{S}_n and $\underline{\mathbf{S}}_n$ differ by $|p - n|$ number of zero eigenvalues; hence, $F^{\underline{\mathbf{S}}_n} = (1 - y_n)I_{[0, \infty)} + y_n F^{\mathbf{S}_n}$, and

$$\underline{m}_n(z) = -\frac{1 - y_n}{z} + y_n m_n(z). \quad (\text{A.13})$$

Thus, equation (A.2) can also be expressed as

$$\begin{aligned} t_n &= y_n - 1 + y_n(z - t_n\sigma_n^2)m_n(z - t_n\sigma_n^2) \\ &= (z - t_n\sigma_n^2)\underline{m}_n(z - t_n\sigma_n^2). \end{aligned}$$

Taking real parts on both sides yields

$$\Re(t_n) = \int \frac{x(u - \Re(t_n)\sigma_n^2) - |z - t_n\sigma_n^2|^2}{|x - z + t_n\sigma_n^2|^2} dF^{\mathbf{S}_n}(x).$$

Solving for $\Re(t_n)$ yields

$$\Re(t_n) = \frac{\int \frac{xu - |z - t_n\sigma_n^2|^2}{|x - z + t_n\sigma_n^2|^2} dF^{\mathbf{S}_n}(x)}{1 + \sigma_n^2 \int \frac{x}{|x - z + t_n\sigma_n^2|^2} dF^{\mathbf{S}_n}(x)} \quad (\text{A.14})$$

Now suppose that $\{t_{n_1} = \Re(t_n)\}$ is not tight; then, with positive probability, there exists a subsequence $\{n_k\}$ such that $|\Re(t_{n_k})| \rightarrow \infty$. By (A.14), we have

$$|\Re(t_{n_k})| \leq \int_{a_{n_k}}^{b_{n_k}} \frac{x|u| + |z - t_{n_k}\sigma_{n_k}^2|^2}{|x - z + t_{n_k}\sigma_{n_k}^2|^2} dF^{\mathbf{S}_{n_k}}(x).$$

However, as k goes to infinity, if $|\Re(t_{n_k})| \rightarrow \infty$, because $\{F^{\mathbf{S}_{n_k}}\}$ is tight and $\sigma_{n_k} \rightarrow \sigma > 0$, one gets that the RHS goes to 1. This contradicts the supposition that $|\Re(t_{n_k})| \rightarrow \infty$.

Next, for any convergent subsequence $\{t_{n_k}\}$ in set \mathcal{D} , by (A.7), for all sufficiently large n_k , we have $v - \Im(t_{n_k})\sigma_{n_k}^2 \geq v/2$. We can then apply the dominated convergence theorem to conclude that the limit point of $\{t_{n_k}\}$ must satisfy equation (A.5). By Lemma A.3, the solution is unique; hence, the whole sequence $\{t_n\}$ converges to the unique solution to equation (A.5). \square

A.2. Some further preliminary results

Let $K^* = \max\{K_1, K_2, K_3\} (= K_3)$ for K_1 , K_2 and K_3 as defined in Lemmas A.1, A.3 and A.4, respectively. Also define $\mathbb{C}^* = \{z \in \mathbb{C}^+ : \Im(z) > K^*\}$. Below we work with $z \in \mathbb{C}^*$.

Let \mathbf{a}_j and \mathbf{e}_j , $j = 1, \dots, n$, be the j th column of \mathbf{A}_n and \mathbf{E}_n , and let $\mathbf{b}_j = \sigma_n \mathbf{e}_j$. Denote $\boldsymbol{\xi}_j = (\mathbf{a}_j + \mathbf{b}_j)/\sqrt{n}$ so that $\mathbf{S}_n = \sum_{j=1}^n \boldsymbol{\xi}_j \boldsymbol{\xi}_j^T$. For any complex

number t_n such that $\Im(z - t_n \sigma_n^2) > 0$, define

$$\begin{aligned} \mathbf{R}_n &= \mathbf{S}_n - (z - t_n \sigma_n^2) \mathbf{I}, & \delta_n &= \frac{\sigma_n^2}{n} \text{tr}(\mathbf{R}_n^{-1}) = y_n \sigma_n^2 \frac{1}{p} \text{tr}(\mathbf{R}_n^{-1}), \\ \mathbf{S}_{nj} &= \mathbf{S}_n - \boldsymbol{\xi}_j \boldsymbol{\xi}_j^T = \sum_{k \neq j} \boldsymbol{\xi}_k \boldsymbol{\xi}_k^T, & \mathbf{R}_{nj} &= \mathbf{S}_{nj} - (z - t_n \sigma_n^2) \mathbf{I}, \\ \mathbf{B}_n &= \frac{1}{1 + \delta_n} \frac{1}{n} \mathbf{A}_n \mathbf{A}_n^T - z \mathbf{I}, \text{ and } \beta_j = \frac{1}{1 + \boldsymbol{\xi}_j^T \mathbf{R}_{nj}^{-1} \boldsymbol{\xi}_j}. \end{aligned} \quad (\text{A.15})$$

According to equation (2.2) in Silverstein and Bai (1995), we have

$$\boldsymbol{\xi}_j^T \mathbf{R}_n^{-1} = \frac{\boldsymbol{\xi}_j^T \mathbf{R}_{nj}^{-1}}{1 + \boldsymbol{\xi}_j^T \mathbf{R}_{nj}^{-1} \boldsymbol{\xi}_j} = \beta_j \boldsymbol{\xi}_j^T \mathbf{R}_{nj}^{-1}. \quad (\text{A.16})$$

Thus, using the identity $\mathbf{A}^{-1} - \mathbf{B}^{-1} = \mathbf{A}^{-1}(\mathbf{B} - \mathbf{A})\mathbf{B}^{-1}$, we obtain

$$\mathbf{R}_n^{-1} = \mathbf{R}_{nj}^{-1} - \mathbf{R}_n^{-1} \boldsymbol{\xi}_j \boldsymbol{\xi}_j^T \mathbf{R}_{nj}^{-1} = \mathbf{R}_{nj}^{-1} - \beta_j \mathbf{R}_{nj}^{-1} \boldsymbol{\xi}_j \boldsymbol{\xi}_j^T \mathbf{R}_{nj}^{-1}. \quad (\text{A.17})$$

Next, we introduce another definition of t_n as the solution to the following equation

$$t_n = -\frac{1}{n} \sum_{j=1}^n \beta_j = -\frac{1}{n} \sum_{j=1}^n \frac{1}{1 + \boldsymbol{\xi}_j^T \mathbf{R}_{nj}^{-1} \boldsymbol{\xi}_j}. \quad (\text{A.18})$$

We claim that the definition of t_n in (A.18) is equivalent to the earlier definition of defining t_n to be the solution to equation (A.2). In fact, write

$$\mathbf{R}_n + z \mathbf{I} = \sum_{j=1}^n \boldsymbol{\xi}_j \boldsymbol{\xi}_j^T + t_n \sigma_n^2 \mathbf{I}.$$

Right-multiplying both sides by \mathbf{R}_n^{-1} and using (A.16) yield

$$\mathbf{I} + z \mathbf{R}_n^{-1} = \sum_{j=1}^n \boldsymbol{\xi}_j \boldsymbol{\xi}_j^T \mathbf{R}_n^{-1} + t_n \sigma_n^2 \mathbf{R}_n^{-1} = \sum_{j=1}^n \frac{\boldsymbol{\xi}_j \boldsymbol{\xi}_j^T \mathbf{R}_{nj}^{-1}}{1 + \boldsymbol{\xi}_j^T \mathbf{R}_{nj}^{-1} \boldsymbol{\xi}_j} + t_n \sigma_n^2 \mathbf{R}_n^{-1}.$$

Taking the trace on both sides and dividing by n , one gets

$$\begin{aligned} y_n + z \frac{1}{n} \text{tr}(\mathbf{R}_n^{-1}) &= 1 - \frac{1}{n} \sum_{j=1}^n \frac{1}{1 + \boldsymbol{\xi}_j^T \mathbf{R}_{nj}^{-1} \boldsymbol{\xi}_j} + t_n \sigma_n^2 \frac{1}{n} \text{tr}(\mathbf{R}_n^{-1}) \\ &= 1 - \frac{1}{n} \sum_{j=1}^n \beta_j + t_n \sigma_n^2 \frac{1}{n} \text{tr}(\mathbf{R}_n^{-1}). \end{aligned} \quad (\text{A.19})$$

This shows that if t_n satisfies (A.18), then t_n satisfies equation (A.2). On the other hand, if t_n satisfies equation (A.2), from (A.19) we have

$$-\frac{1}{n} \sum_{j=1}^n \beta_j = y_n - 1 + (z - t_n \sigma_n^2) \frac{1}{n} \text{tr}(\mathbf{R}_n)^{-1} = t_n,$$

namely, t_n satisfies (A.18).

We proceed to analyze the difference in (A.1). Because

$$\begin{aligned} \mathbf{S}_n - \frac{1}{1+\delta_n} \frac{1}{n} \mathbf{A}_n \mathbf{A}_n^T &= \frac{1}{n} \sum_{j=1}^n (\mathbf{a}_j + \mathbf{b}_j)(\mathbf{a}_j + \mathbf{b}_j)^T - \frac{1}{1+\delta_n} \frac{1}{n} \sum_{j=1}^n \mathbf{a}_j \mathbf{a}_j^T \\ &= \frac{1}{n} \sum_{j=1}^n \left(\frac{\delta_n}{1+\delta_n} \mathbf{a}_j \mathbf{a}_j^T + \mathbf{a}_j \mathbf{b}_j^T + \mathbf{b}_j \mathbf{a}_j^T + \mathbf{b}_j \mathbf{b}_j^T \right), \end{aligned}$$

and recall that $\mathbf{R}_n = \mathbf{S}_n - (z - t_n \sigma_n^2) \mathbf{I}$ and $\mathbf{B}_n = \frac{1}{1+\delta_n} \frac{1}{n} \mathbf{A}_n \mathbf{A}_n^T - z \mathbf{I}$, we have

$$\begin{aligned} \Delta &:= \frac{1}{p} \text{tr}(\mathbf{B}_n^{-1} - \mathbf{R}_n^{-1}) \\ &= \frac{1}{p} \text{tr} \left(\mathbf{B}_n^{-1} \left(\mathbf{S}_n - \frac{1}{1+\delta_n} \frac{1}{n} \mathbf{A}_n \mathbf{A}_n^T + t_n \sigma_n^2 \mathbf{I} \right) \mathbf{R}_n^{-1} \right) \\ &= \frac{1}{np} \sum_{j=1}^n \left\{ \frac{\delta_n}{1+\delta_n} \mathbf{a}_j^T \mathbf{R}_n^{-1} \mathbf{B}_n^{-1} \mathbf{a}_j + \mathbf{b}_j^T \mathbf{R}_n^{-1} \mathbf{B}_n^{-1} \mathbf{a}_j + \mathbf{a}_j^T \mathbf{R}_n^{-1} \mathbf{B}_n^{-1} \mathbf{b}_j + \mathbf{b}_j^T \mathbf{R}_n^{-1} \mathbf{B}_n^{-1} \mathbf{b}_j \right\} \\ &\quad + \frac{t_n \sigma_n^2}{p} \text{tr}(\mathbf{R}_n^{-1} \mathbf{B}_n^{-1}). \end{aligned}$$

Using (A.17), we obtain

$$\begin{aligned} \Delta &= \frac{1}{np} \sum_{j=1}^n \left[\frac{\delta_n}{1+\delta_n} \mathbf{a}_j^T \mathbf{R}_{nj}^{-1} \mathbf{B}_n^{-1} \mathbf{a}_j - \frac{\delta_n}{1+\delta_n} \beta_j \mathbf{a}_j^T \mathbf{R}_{nj}^{-1} \boldsymbol{\xi}_j \boldsymbol{\xi}_j^T \mathbf{R}_{nj}^{-1} \mathbf{B}_n^{-1} \mathbf{a}_j \right. \\ &\quad + \mathbf{b}_j^T \mathbf{R}_{nj}^{-1} \mathbf{B}_n^{-1} \mathbf{a}_j - \beta_j \mathbf{b}_j^T \mathbf{R}_{nj}^{-1} \boldsymbol{\xi}_j \boldsymbol{\xi}_j^T \mathbf{R}_{nj}^{-1} \mathbf{B}_n^{-1} \mathbf{a}_j \\ &\quad + \mathbf{a}_j^T \mathbf{R}_{nj}^{-1} \mathbf{B}_n^{-1} \mathbf{b}_j - \beta_j \mathbf{a}_j^T \mathbf{R}_{nj}^{-1} \boldsymbol{\xi}_j \boldsymbol{\xi}_j^T \mathbf{R}_{nj}^{-1} \mathbf{B}_n^{-1} \mathbf{b}_j \\ &\quad \left. + \mathbf{b}_j^T \mathbf{R}_{nj}^{-1} \mathbf{B}_n^{-1} \mathbf{b}_j - \beta_j \mathbf{b}_j^T \mathbf{R}_{nj}^{-1} \boldsymbol{\xi}_j \boldsymbol{\xi}_j^T \mathbf{R}_{nj}^{-1} \mathbf{B}_n^{-1} \mathbf{b}_j \right] \\ &\quad + \frac{t_n \sigma_n^2}{p} \text{tr}(\mathbf{R}_n^{-1} \mathbf{B}_n^{-1}). \end{aligned}$$

Define

$$\begin{aligned} \rho_j &= \frac{1}{n} \mathbf{a}_j^T \mathbf{R}_{nj}^{-1} \mathbf{a}_j, & \hat{\rho}_j &= \frac{1}{n} \mathbf{a}_j^T \mathbf{R}_{nj}^{-1} \mathbf{B}_n^{-1} \mathbf{a}_j, \\ w_j &= \frac{1}{n} \mathbf{b}_j^T \mathbf{R}_{nj}^{-1} \mathbf{b}_j, & \hat{w}_j &= \frac{1}{n} \mathbf{b}_j^T \mathbf{R}_{nj}^{-1} \mathbf{B}_n^{-1} \mathbf{b}_j, \\ \eta_j &= \frac{1}{n} \mathbf{a}_j^T \mathbf{R}_{nj}^{-1} \mathbf{b}_j, & \hat{\eta}_j &= \frac{1}{n} \mathbf{a}_j^T \mathbf{R}_{nj}^{-1} \mathbf{B}_n^{-1} \mathbf{b}_j, \\ \gamma_j &= \frac{1}{n} \mathbf{b}_j^T \mathbf{R}_{nj}^{-1} \mathbf{a}_j, & \hat{\gamma}_j &= \frac{1}{n} \mathbf{b}_j^T \mathbf{R}_{nj}^{-1} \mathbf{B}_n^{-1} \mathbf{a}_j. \end{aligned} \tag{A.20}$$

Certainly, $\eta_j = \gamma_j$, but introducing γ_j makes the computations below more clear.

Recall that $\boldsymbol{\xi}_j = (\mathbf{a}_j + \mathbf{b}_j)/\sqrt{n}$, and so $\beta_j^{-1} = 1 + \rho_j + w_j + \eta_j + \gamma_j$. We can then rewrite Δ as

$$\begin{aligned}
\Delta &= \frac{1}{p} \sum_{j=1}^n \beta_j \left(\frac{\delta_n}{1 + \delta_n} \hat{\rho}_j (1 + \rho_j + \eta_j + \gamma_j + w_j) - \frac{\delta_n}{1 + \delta_n} (\rho_j + \eta_j) (\hat{\rho}_j + \hat{\gamma}_j) \right. \\
&\quad + \hat{\gamma}_j (1 + \rho_j + \eta_j + \gamma_j + w_j) - (\gamma_j + w_j) (\hat{\gamma}_j + \hat{\rho}_j) \\
&\quad + \hat{\eta}_j (1 + \rho_j + \eta_j + \gamma_j + w_j) - (\rho_j + \eta_j) (\hat{\eta}_j + \hat{w}_j) \\
&\quad \left. + \hat{w}_j (1 + \rho_j + \eta_j + \gamma_j + w_j) - (\gamma_j + w_j) (\hat{\eta}_j + \hat{w}_j) \right) \\
&\quad + \frac{t_n \sigma_n^2}{p} \text{tr}(\mathbf{R}_n^{-1} \mathbf{B}_n^{-1}) \\
&= \frac{1}{p} \sum_{j=1}^n \beta_j \left(\frac{1}{1 + \delta_n} \hat{\rho}_j (\delta_n - \gamma_j - w_j) + \hat{\gamma}_j \left(1 + \frac{1}{1 + \delta_n} (\rho_j + \eta_j) \right) + \hat{\eta}_j + \hat{w}_j \right) \\
&\quad + \frac{t_n \sigma_n^2}{p} \text{tr}(\mathbf{R}_n^{-1} \mathbf{B}_n^{-1}) \\
&:= \Delta_1 + \Delta_2 + \Delta_3,
\end{aligned}$$

where

$$\begin{aligned}
\Delta_1 &= \frac{1}{p(1 + \delta_n)} \sum_{j=1}^n \beta_j \hat{\rho}_j (\delta_n - w_j) - \frac{1}{p(1 + \delta_n)} \sum_{j=1}^n \beta_j \hat{\rho}_j \gamma_j, \\
\Delta_2 &= \frac{1}{p} \sum_{j=1}^n \beta_j \hat{\gamma}_j \left(1 + \frac{1}{1 + \delta_n} (\rho_j + \eta_j) \right) + \frac{1}{p} \sum_{j=1}^n \beta_j \hat{\eta}_j, \quad \text{and} \\
\Delta_3 &= \frac{1}{p} \sum_{j=1}^n \beta_j \left(\hat{w}_j - \frac{\sigma_n^2}{n} \text{tr}(\mathbf{R}_n^{-1} \mathbf{B}_n^{-1}) \right),
\end{aligned} \tag{A.21}$$

and in the last equality we used the equivalent definition (A.18) of t_n .

Lemma A.6. Suppose that t_n solves equation (A.2) for $z = u + iv \in \mathbb{C}^*$; then,

- (i) for all $j = 1, \dots, n$, $|\beta_j|$ is bounded by $\frac{|z - t_n \sigma_n^2|}{v - \Im(t_n) \sigma_n^2}$;
- (ii) $\|\mathbf{B}_n^{-1}\|$ is bounded by v^{-1} ;
- (iii) the random variables ϖ_j satisfy

$$\max_{1 \leq j \leq n} E|\varpi_j|^4 \leq \frac{C(\log n)^6}{n^2(v - t_{n2} \sigma_n^2)^4},$$

where ϖ_j can be any of η_j , $\hat{\eta}_j$, γ_j and $\hat{\gamma}_j$ defined in (A.20), and C is a constant independent of n ;

(iv) the random variables w_j and \hat{w}_j satisfy

$$\begin{aligned} \max_{1 \leq j \leq n} E \left| w_j - \frac{\sigma_n^2}{n} \text{tr}(\mathbf{R}_n^{-1}) \right|^4 &\leq \frac{C(\log n)^8}{n^2(v - t_{n2}\sigma_n^2)^4}, \\ \max_{1 \leq j \leq n} E \left| \hat{w}_j - \frac{\sigma_n^2}{n} \text{tr}(\mathbf{R}_n^{-1} \mathbf{B}_n^{-1}) \right|^4 &\leq \frac{C(\log n)^8}{n^2 v^4 (v - t_{n2}\sigma_n^2)^4}. \end{aligned}$$

Proof. We first prove (i). Write $t_n = t_{n1} + it_{n2}$. Note that

$$\begin{aligned} &\Im \left\{ (z - t_n \sigma_n^2) \boldsymbol{\xi}_j^T \mathbf{R}_{nj}^{-1} \boldsymbol{\xi}_j \right\} \\ = &\Im \left\{ \boldsymbol{\xi}_j^T \left(\frac{1}{z - t_n \sigma_n^2} \mathbf{S}_{nj} - \mathbf{I} \right)^{-1} \boldsymbol{\xi}_j \right\} \\ = &\frac{1}{2i} \boldsymbol{\xi}_j^T \left[\left(\frac{1}{z - t_n \sigma_n^2} \mathbf{S}_{nj} - \mathbf{I} \right)^{-1} - \left(\frac{1}{\bar{z} - t_n \sigma_n^2} \mathbf{S}_{nj} - \mathbf{I} \right)^{-1} \right] \boldsymbol{\xi}_j \\ = &\frac{v - t_{n2}\sigma_n^2}{|z - t_n \sigma_n^2|^2} \boldsymbol{\xi}_j^T \left(\frac{1}{z - t_n \sigma_n^2} \mathbf{S}_{nj} - \mathbf{I} \right)^{-1} \mathbf{S}_{nj} \left(\frac{1}{\bar{z} - t_n \sigma_n^2} \mathbf{S}_{nj} - \mathbf{I} \right)^{-1} \boldsymbol{\xi}_j \\ \geq &0, \end{aligned}$$

where the last inequality is because of Lemma A.5. Therefore,

$$\begin{aligned} |\beta_j| &= \frac{|z - t_n \sigma_n^2|}{|(z - t_n \sigma_n^2)(1 + \boldsymbol{\xi}_j^T \mathbf{R}_{nj}^{-1} \boldsymbol{\xi}_j)|} \\ &\leq \frac{|z - t_n \sigma_n^2|}{|\Im \{ (z - t_n \sigma_n^2)(1 + \boldsymbol{\xi}_j^T \mathbf{R}_{nj}^{-1} \boldsymbol{\xi}_j) \}|} \\ &\leq \frac{|z - t_n \sigma_n^2|}{v - t_{n2}\sigma_n^2}. \end{aligned}$$

As to (ii), note that any eigenvalue of $\mathbf{B}_n = \frac{1}{n(1 + \delta_n)} \mathbf{A}_n \mathbf{A}_n^T - z \mathbf{I}$ can be expressed as $\lambda^B = \frac{1}{1 + \delta_n} \lambda - z$, where λ is an eigenvalue of $\frac{1}{n} \mathbf{A}_n \mathbf{A}_n^T$. We have

$$|\lambda^B| \geq |\Im(\lambda^B)| = \left| \frac{\Im(\delta_n)}{|1 + \delta_n|^2} \lambda + v \right| \geq v,$$

where the last step follows from the fact that $\Im(\delta_n) = y_n \sigma_n^2 \Im(m_n(z - t_n \sigma_n^2)) > 0$, because of Lemma A.5.

Now we prove (iii). We shall only establish the inequality for $\eta_j (= \gamma_j)$; the other two variables $\hat{\eta}_j$ and $\hat{\gamma}_j$ can be handled in a similar way by using (ii).

Because for any Hermitian matrix \mathbf{A} and $z \in \mathbb{C}^+$, $\|(\mathbf{A} - z \mathbf{I})^{-1}\| \leq 1/\Im(z)$, we have by Lemma A.5 that

$$\|\mathbf{R}_n^{-1}\| \leq \frac{1}{(v - t_{n2}\sigma_n^2)}, \quad \text{and} \quad \max_{1 \leq j \leq n} \|\mathbf{R}_{nj}^{-1}\| \leq \frac{1}{(v - t_{n2}\sigma_n^2)}. \quad (\text{A.22})$$

Recall that $\mathbf{b}_j = \sigma_n \boldsymbol{\epsilon}_j$, and $\boldsymbol{\epsilon}_j$ satisfies $E(\boldsymbol{\epsilon}_j \boldsymbol{\epsilon}_j^T) = \mathbf{I}$. The strengthened assumption (E.iii) implies that $|\mathbf{a}_j| \leq C\sqrt{n \log n}$. Note also that $\boldsymbol{\epsilon}_j$ is independent of \mathbf{R}_{nj}^{-1} and \mathbf{a}_j . Moreover, using Lemma E.1 in Appendix E, Assumption (E.i) and (A.22), we get

$$\begin{aligned}
 E|\eta_j|^4 &= \frac{1}{n^4} E|\mathbf{a}_j^T \mathbf{R}_{nj}^{-1} \mathbf{b}_j|^4 = \frac{\sigma_n^4}{n^4} E|\mathbf{a}_j^T \mathbf{R}_{nj}^{-1} \boldsymbol{\epsilon}_j|^4 \\
 &= \frac{\sigma_n^4}{n^4} E(\boldsymbol{\epsilon}_j^T \bar{\mathbf{R}}_{nj}^{-1} \mathbf{a}_j \mathbf{a}_j^T \mathbf{R}_{nj}^{-1} \boldsymbol{\epsilon}_j)^2 \\
 &\leq \frac{2\sigma_n^4}{n^4} (E|\boldsymbol{\epsilon}_j^T \bar{\mathbf{R}}_{nj}^{-1} \mathbf{a}_j \mathbf{a}_j^T \mathbf{R}_{nj}^{-1} \boldsymbol{\epsilon}_j - \mathbf{a}_j^T \mathbf{R}_{nj}^{-1} \bar{\mathbf{R}}_{nj}^{-1} \mathbf{a}_j|^2 + E(\mathbf{a}_j^T \mathbf{R}_{nj}^{-1} \bar{\mathbf{R}}_{nj}^{-1} \mathbf{a}_j)^2) \\
 &\leq \frac{C}{n^4} E|\epsilon_{11}|^4 \times E(\mathbf{a}_j^T \mathbf{R}_{nj}^{-1} \bar{\mathbf{R}}_{nj}^{-1} \mathbf{a}_j)^2 \\
 &\leq \frac{C(\log n)^6}{n^2(v - t_{n2}\sigma_n^2)^4}.
 \end{aligned}$$

Finally, we prove (iv). Using (E.i), (A.22), (ii), Lemma E.1 and Lemma 2.6 in Silverstein and Bai (1995), we obtain

$$\begin{aligned}
 &E \left| w_j - \frac{\sigma_n^2}{n} \text{tr}(\mathbf{R}_n^{-1}) \right|^4 \\
 &\leq C \left(E \left| \frac{\sigma_n^2}{n} \boldsymbol{\epsilon}_j^T \mathbf{R}_{nj}^{-1} \boldsymbol{\epsilon}_j - \frac{\sigma_n^2}{n} \text{tr}(\mathbf{R}_{nj}^{-1}) \right|^4 + E \left| \frac{\sigma_n^2}{n} \text{tr}(\mathbf{R}_{nj}^{-1} - \mathbf{R}_n^{-1}) \right|^4 \right) \\
 &\leq \frac{C}{n^4} \left| E((\log n)^4 \text{tr}(\mathbf{R}_{nj}^{-1} \bar{\mathbf{R}}_{nj}^{-1}))^2 + (\log n)^8 E \text{tr}(\mathbf{R}_{nj}^{-1} \bar{\mathbf{R}}_{nj}^{-1})^2 \right| + \frac{C}{n^4(v - t_{n2}\sigma_n^2)^4} \\
 &\leq \frac{C(\log n)^8}{n^2(v - t_{n2}\sigma_n^2)^4}.
 \end{aligned}$$

The result for \hat{w}_j can be proved similarly. \square

A.3. Proof of Proposition A.1

Proof of Proposition A.1. First, the existence and uniqueness of t_n have been established in Lemma A.1.

Next, to show (A.1), we recall the $\Delta_j, j = 1, 2, 3$ defined in (A.21). The proof is completed if we show $\Delta_j \rightarrow 0$ almost surely for all $j = 1, 2, 3$.

By (A.22), (E.iii) and Lemma A.6 (ii), there exists a constant C such that

$$\max_{j=1, \dots, n} |\rho_j| \leq \frac{C \log(n)}{v - t_{n2}\sigma_n^2}, \quad \text{and} \quad \max_{j=1, \dots, n} |\hat{\rho}_j| \leq \frac{C \log(n)}{v(v - t_{n2}\sigma_n^2)}. \quad (\text{A.23})$$

Moreover, by Lemmas A.5 and A.2 and the convergence of $\{F^{\mathbf{S}_n}\}$, we have that as $p \rightarrow \infty$,

$$\delta_n = y_n \sigma_n m_n(z - t_n \sigma_n^2) \rightarrow \delta = \delta(z) = y \sigma^2 m(z - t \sigma^2), \quad (\text{A.24})$$

and $\Im(\delta) > 0$. In particular, for all sufficiently large n , we have

$$\frac{1}{|1 + \delta_n|} \leq \frac{2}{\liminf_n \Im(\delta_n)} < \infty. \quad (\text{A.25})$$

We now show that $\Delta_3 \rightarrow 0$ almost surely. Using Markov's inequality and Hölder's inequality, for any $\varepsilon > 0$, we have

$$\begin{aligned} \mathbb{P}(|\Delta_3| \geq \varepsilon) &\leq \frac{1}{\varepsilon^4} E \left| \frac{1}{p} \sum_{j=1}^n \beta_j \left(\hat{w}_j - \frac{\sigma_n^2}{n} \text{tr}(\mathbf{R}_n^{-1} \mathbf{B}_n^{-1}) \right) \right|^4 \\ &\leq \frac{n^3}{p^4 \varepsilon^4} \sum_{j=1}^n E |\beta_j|^4 \left| \hat{w}_j - \frac{\sigma_n^2}{n} \text{tr}(\mathbf{R}_n^{-1} \mathbf{B}_n^{-1}) \right|^4 \\ &\leq \frac{C(\log n)^8}{n^2 \varepsilon^4 v^4 (v - t_{n2} \sigma_n^2)^8} \cdot |z - t_n \sigma_n^2|^4, \end{aligned}$$

where the last step follows from Lemma A.6 (i) and (iv). Thus, $\Delta_3 \rightarrow 0$ almost surely by Lemmas A.5 and A.2 and the Borel-Cantelli Lemma.

Similarly we can prove that $\Delta_j \rightarrow 0$ almost surely for $j = 1, 2$ by using Lemmas A.5, A.2, and A.6 and inequalities (A.23) and (A.25). \square

A.4. Proof of Theorem 2.1

Proof of Theorem 2.1. We first show that equation (1.1) in Dozier and Silverstein (2007a) can be derived from Proposition A.1. For any fixed $z \in \mathbb{C}^*$, by Proposition A.1, Lemmas A.5, A.2, and Lemma A.6 (ii), and the dominated convergence theorem, we obtain

$$m(z - t\sigma^2) = \int \frac{1}{(1 + \delta)^{-1}x - z} dF^{\mathcal{A}}(x), \quad (\text{A.26})$$

where t is the unique solution to equation (A.5), $\delta = y\sigma^2 m(z - t\sigma^2)$, and $m(\cdot)$ is the Stieltjes transform of the probability distribution F . Moreover, if we let $\gamma(z) = z - t(z)\sigma^2$, then by (A.5) and (A.24), we have

$$t = y - 1 + y\gamma m(\gamma), \quad \delta = y\sigma^2 m(\gamma),$$

and

$$z = \gamma + t\sigma^2 = \gamma + \gamma y\sigma^2 m(\gamma) + \sigma^2(y - 1).$$

Substituting the expressions of t , δ , and z in terms of γ into equation (A.26) yields

$$m(\gamma) = \int \frac{dF^{\mathcal{A}}(x)}{\frac{x}{1 + y\sigma^2 m(\gamma)} - \gamma(1 + y\sigma^2 m(\gamma)) - \sigma^2(y - 1)}, \quad (\text{A.27})$$

where $\gamma \in \mathbb{C}_\gamma := \{\gamma = z - t(z)\sigma^2 : z \in \mathbb{C}^*\}$. Note further that by Lemma A.4, $\gamma(z)$ is analytic on \mathbb{C}^* . It then follows from the uniqueness of analytic continuation that equation (A.27) holds for every $\gamma \in \mathbb{C}^+$ – in other words, equation (1.1) in Dozier and Silverstein (2007a) holds.

In the following, we show that equation (2.5) in Theorem 2.1 holds.

For any $z \in \mathbb{C}^*$, denote $\alpha(z) = z(1 + \delta(z))$, where $\delta(z) = y\sigma^2 m(\gamma)$ and $\gamma = z - t\sigma^2$. We further define

$$d(\gamma) = 1 + y\sigma^2 m(\gamma) (= 1 + \delta(z)), \quad \text{and} \quad g(\alpha) = 1 - y\sigma^2 m_{\mathcal{A}}(\alpha).$$

We show the following facts:

- (F.i) $g(\alpha) = 1/d(\gamma)$,
- (F.ii) $\alpha = \gamma d^2(\gamma) + \sigma^2(y-1)d(\gamma)$, or equivalently,
 $\gamma = \alpha g^2(\alpha) - \sigma^2(y-1)g(\alpha)$.

In fact, we can rewrite equation (A.26) as

$$m_{\mathcal{A}}(\alpha) = (1 + \delta)^{-1} m(\gamma).$$

Noting that $\delta = y\sigma^2 m(\gamma)$, we have

$$y\sigma^2 m_{\mathcal{A}}(\alpha) = \frac{\delta}{1 + \delta}, \quad \text{and hence} \quad g(\alpha) = \frac{1}{1 + \delta} = \frac{1}{d(\gamma)},$$

namely, (F.i) holds. In addition, $y\sigma^2 m_{\mathcal{A}}(\alpha) = 1 - 1/(1 + \delta)$ implies $\alpha \in \mathbb{C}^+$ because $\delta = y\sigma^2 m(z - t\sigma^2) \in \mathbb{C}^+$ by Lemma A.2.

We now show (F.ii). Let $\beta = t\sigma^2(1 + \delta)$. Then

$$\gamma = z - t\sigma^2 = \frac{\alpha - \beta}{1 + \delta}. \tag{A.28}$$

By substituting (A.28) and $\delta = y\sigma^2 m(\gamma)$ into equation (A.5), we obtain

$$\frac{\beta}{\sigma^2(1 + \delta)} = y - 1 + \frac{\delta(\alpha - \beta)}{\sigma^2(1 + \delta)}.$$

That is,

$$\beta = \sigma^2(y - 1) + \frac{\delta}{1 + \delta} \alpha.$$

Therefore,

$$\begin{aligned} \gamma &= \frac{\alpha - \beta}{1 + \delta} = \frac{\alpha}{(1 + \delta)^2} - \frac{\sigma^2(y - 1)}{1 + \delta} \\ &= \alpha g^2(\alpha) - \sigma^2(y - 1)g(\alpha), \end{aligned}$$

namely, (F.ii) holds.

Next, by (A.27) and the definitions of α and $d(\gamma)$ and (F.ii), we have

$$m(\gamma) = d(\gamma) \int \frac{1}{x - \alpha} dF^{\mathcal{A}}(x).$$

Using the facts (F.i) and (F.ii), we obtain

$$\begin{aligned} m_{\mathcal{A}}(\alpha) &= \int \frac{dF^{\mathcal{A}}(x)}{x - \alpha} = \frac{1}{d(\gamma)} \int \frac{1}{\tau - \gamma} dF(\tau) \\ &= \int \frac{g(\alpha)}{\tau - \alpha g^2(\alpha) + \sigma^2(y - 1)g(\alpha)} dF(\tau) \\ &= \int \frac{1}{\frac{\tau}{g(\alpha)} - \alpha g(\alpha) + \sigma^2(y - 1)} dF(\tau). \end{aligned} \quad (\text{A.29})$$

By plugging in the expression of $g(\alpha)$, we see that for all $\alpha = \alpha(z) = z(1 + \delta(z))$, $m_{\mathcal{A}}(\alpha)$ satisfies

$$m_{\mathcal{A}}(\alpha) = \int \frac{dF(\tau)}{\frac{\tau}{1 - y\sigma^2 m_{\mathcal{A}}(\alpha)} - \alpha(1 - y\sigma^2 m_{\mathcal{A}}(\alpha)) + \sigma^2(y - 1)}.$$

It follows from the uniqueness of analytic continuation that the above equation holds for all $\alpha \in \mathbb{C}^+$ such that $m_{\mathcal{A}}(\alpha) \in D_{\mathcal{A}}$.

It remains to show that the solution to equation (2.5) is unique in $D_{\mathcal{A}}$. Suppose that $m \in D_{\mathcal{A}}$ satisfies equation (2.5). Define

$$\gamma = z(1 - y\sigma^2 m)^2 - \sigma^2(y - 1)(1 - y\sigma^2 m) \in \mathbb{C}^+. \quad (\text{A.30})$$

By (2.5) and (A.30), we have $m = (1 - y\sigma^2 m)m(\gamma)$. Hence,

$$m = \frac{m(\gamma)}{1 + y\sigma^2 m(\gamma)}, \quad \text{and} \quad m(\gamma) = \frac{m}{1 - y\sigma^2 m}. \quad (\text{A.31})$$

The second identity implies that

$$1 + y\sigma^2 m(\gamma) = \frac{1}{1 - y\sigma^2 m}. \quad (\text{A.32})$$

Using (A.30) and (A.32), we have

$$\begin{aligned} z &= \frac{\gamma}{(1 - y\sigma^2 m)^2} + \frac{\sigma^2(y - 1)}{1 - y\sigma^2 m} \\ &= \gamma(1 + y\sigma^2 m(\gamma))^2 + \sigma^2(y - 1)(1 + y\sigma^2 m(\gamma)). \end{aligned} \quad (\text{A.33})$$

Because the Stieltjes transform $m(\gamma)$ is uniquely determined by equation (A.27), we obtain

$$\begin{aligned} m(\gamma) &= \int \frac{dF^{\mathcal{A}}(x)}{\frac{x}{1 + y\sigma^2 m(\gamma)} - \gamma(1 + y\sigma^2 m(\gamma)) - \sigma^2(y - 1)} \\ &= (1 + y\sigma^2 m(\gamma)) \cdot m_{\mathcal{A}}(z). \end{aligned}$$

It then follows from the first identity in (A.31) that $m = m_{\mathcal{A}}(z)$ – in other words, $m_{\mathcal{A}}(z)$ is the unique solution in $D_{\mathcal{A}}$. \square

Appendix B: Proof of Theorem 2.2

Proof of Theorem 2.2. The convergence of F^{ICV} follows easily from Assumption (C.iii) and the fact that

$$F^{\text{ICV}}(x) = F^{\tilde{\Sigma}} \left(\frac{x}{\int_0^1 \gamma_t^2 dt} \right) \quad \text{for all } x \geq 0.$$

Next, by Theorem 3.2 in Dozier and Silverstein (2007b), the assumption that F has a bounded support implies that H also has a bounded support. Thus, Assumption (A.iii') in Zheng and Li (2011), which requires that H has a finite second moment, is satisfied.

We proceed to show the convergence of \mathcal{A}_m . As discussed in Section 2.2.2, if the diffusion process (\mathbf{X}_t) belongs to Class \mathcal{C} , the drift process $\boldsymbol{\mu}_t \equiv 0$, and (γ_t) is independent of (\mathbf{W}_t) , then, conditional on $\{\gamma_t\}$, we have

$$\Delta \bar{\mathbf{X}}_{2i} \stackrel{\mathcal{D}}{=} \sqrt{w_i} \tilde{\Sigma}^{1/2} \mathbf{Z}_i, \quad (\text{B.1})$$

where w_i is as in (2.12) and is independent of \mathbf{Z}_i , and $\mathbf{Z}_i = (Z_i^1, \dots, Z_i^p)^T$ consists of independent standard normals. Hence, \mathcal{A}_m has the same distribution as $\tilde{\mathcal{A}}_m$ defined as

$$\tilde{\mathcal{A}}_m = 3 \sum_{i=1}^m w_i \tilde{\Sigma}^{1/2} \mathbf{Z}_i \mathbf{Z}_i^T \tilde{\Sigma}^{1/2}. \quad (\text{B.2})$$

Claim 1. Without loss of generality, we can assume that the drift process $\boldsymbol{\mu}_t \equiv 0$ and (γ_t) is independent of (\mathbf{W}_t) .

First, whether the drift term $(\boldsymbol{\mu}_t)$ vanishes or not does not affect the LSD of \mathcal{A}_m . To see this, note that $\Delta \bar{\mathbf{X}}_{2i} = \tilde{\mathbf{V}}_i + \tilde{\mathbf{Z}}_i$, where

$$\tilde{\mathbf{V}}_i = \sum_{|j| < k} \left(1 - \frac{|j|}{k} \right) \int_{((2i-1)k+j-1)/n}^{((2i-1)k+j)/n} \boldsymbol{\mu}_t dt, \quad (\text{B.3})$$

and

$$\tilde{\mathbf{Z}}_i = \boldsymbol{\Lambda} \cdot \sum_{|j| < k} \left(1 - \frac{|j|}{k} \right) \int_{((2i-1)k+j-1)/n}^{((2i-1)k+j)/n} \gamma_t d\mathbf{W}_t. \quad (\text{B.4})$$

Because all the entries of $\tilde{\mathbf{V}}_i$ are of order $O(k/n) = o(1/\sqrt{p})$, by Lemma E.2 in Appendix E, \mathcal{A}_m and $3 \sum_{i=1}^m \tilde{\mathbf{Z}}_i (\tilde{\mathbf{Z}}_i)^T$ have the same LSD.

Next, by the same argument as in the beginning of the Proof of Theorem 1 in [Zheng and Li \(2011\)](#), we can assume without loss of generality that (γ_t) is independent of (\mathbf{W}_t) . It follows that \mathcal{A}_m and $\tilde{\mathcal{A}}_m$ have the same LSD.

Claim 2. $\max_{i,n}(mw_i)$ is bounded, and there exists a piecewise continuous process (w_s) with finitely many jumps such that

$$\lim_{n \rightarrow \infty} \sum_{i=1}^m \int_{((2i-2)k)/n}^{2ik/n} |3mw_i - w_s| ds = 0. \quad (\text{B.5})$$

Using the boundedness of (γ_t) assumed in [\(C.vi\)](#) and $k = \lfloor \theta \sqrt{n} \rfloor$, one can easily show that $\max_{i,n}(mw_i)$ is bounded.

Next we show that [\(B.5\)](#) is satisfied for $w_s = (\gamma_s^*)^2$. Define

$$w_i^* = \sum_{|j| < k} \left(1 - \frac{|j|}{k}\right)^2 \int_{((2i-1)k+j-1)/n}^{((2i-1)k+j)/n} (\gamma_t^*)^2 dt.$$

Suppose that (γ_t^*) has J jumps for $J \geq 1$. For each $j = 1, \dots, J$, there exists an ℓ_j such that the j th jump falls in the interval $[(2\ell_j - 2)k/n, (2\ell_j k)/n]$. Then

$$\begin{aligned} & \sum_{i=1}^m \int_{((2i-2)k)/n}^{2ik/n} |3mw_i - w_s| ds \\ &= \sum_{\ell_j \in \{\ell_1, \dots, \ell_J\}} \int_{((2\ell_j-2)k)/n}^{2\ell_j k/n} |3mw_{\ell_j} - w_s| ds \\ & \quad + \sum_{i \notin \{\ell_1, \dots, \ell_J\}} \int_{((2i-2)k)/n}^{2ik/n} |3mw_i - w_s| ds \\ &:= \Delta_1 + \Delta_2. \end{aligned}$$

Because (mw_{ℓ_j}) and $|\gamma_s^*|$ are both bounded, for any $\varepsilon > 0$ and for any sufficiently large n , we have

$$|\Delta_1| \leq \frac{2k}{n} \cdot JC < \varepsilon.$$

For the second term Δ_2 , because (γ_t^*) is continuous in $[(2i-2)k/n, (2ik)/n]$ when $i \notin \{\ell_1, \dots, \ell_J\}$, and by [\(C.vi\)](#), (γ_t) uniformly converges to (γ_t^*) , for any $\varepsilon > 0$ and for sufficiently large n, p , we have

$$|\gamma_t^* - \gamma_{(2i-2)k/n}^*| < \varepsilon \text{ for all } t \in \left[\frac{(2i-2)k}{n}, \frac{2ik}{n} \right], \text{ and } |\gamma_t - \gamma_t^*| < \varepsilon \text{ for all } t.$$

Moreover, because $|\gamma_t| \leq C_2$, for all large n , we have

$$\begin{aligned}
|\Delta_2| &\leq \sum_i \int_{(2i-2)k/n}^{2ik/n} |3mw_i - 3mw_i^*| ds \\
&+ \sum_i \int_{(2i-2)k/n}^{2ik/n} \left| 3mw_i^* - \frac{3m}{n} \left(\gamma_{(2i-2)k/n}^* \right)^2 \cdot \sum_{|j|<k} \left(1 - \frac{|j|}{k} \right)^2 \right| ds \\
&+ \sum_i \int_{(2i-2)k/n}^{2ik/n} \left| \frac{3m}{n} \left(\gamma_{(2i-2)k/n}^* \right)^2 \cdot \sum_{|j|<k} \left(1 - \frac{|j|}{k} \right)^2 - (\gamma_s^*)^2 \right| ds \\
&\leq 3m^2 \cdot \frac{2k}{n} \cdot \frac{2k(k+1)(2k+1)}{6k^2} \cdot \frac{(2C_2\varepsilon)}{n} \\
&+ 3m^2 \cdot \frac{2k}{n} \cdot \frac{2k(k+1)(2k+1)}{6k^2} \cdot \frac{(2C_2\varepsilon)}{n} \\
&+ \sum_i \int_{(2i-2)k/n}^{2ik/n} \left(\gamma_{(2i-2)k/n}^* \right)^2 ds \cdot \left| \frac{3m}{n} \cdot \sum_{|j|<k} \left(1 - \frac{|j|}{k} \right)^2 - 1 \right| \\
&+ \sum_i \int_{(2i-2)k/n}^{2ik/n} \left| \left(\gamma_{(2i-2)k/n}^* \right)^2 - (\gamma_s^*)^2 \right| ds \\
&\leq C\varepsilon.
\end{aligned}$$

This completes the proof of (B.5).

Finally, because $F^{\mathfrak{S}} \rightarrow \check{H}$ and $\check{H}(x/\zeta) = H(x)$ for $x \geq 0$, using Claim 2 and applying Theorem 1 in [Zheng and Li \(2011\)](#), we conclude that the ESD of \mathcal{A}_m converges to $F^{\mathcal{A}}$, whose Stieltjes transform satisfies

$$\begin{aligned}
m_{\mathcal{A}}(z) &= -\frac{1}{z} \int \frac{1}{\tau M(z) + 1} d\check{H}(\tau) \\
&= -\frac{1}{z} \int \frac{\zeta}{\tau M(z) + \zeta} dH(\tau),
\end{aligned} \tag{B.6}$$

where $M(z)$, together with another function $\tilde{m}(z)$, uniquely solve the following equations in $\mathbb{C}^+ \times \mathbb{C}^+$

$$\begin{cases} M(z) &= -\frac{1}{z} \int_0^1 \frac{w_s}{1 + y\tilde{m}(z)w_s} ds, \\ \tilde{m}(z) &= -\frac{1}{z} \int \frac{\tau}{\tau M(z) + 1} d\check{H}(\tau) = -\frac{1}{z} \int \frac{\tau}{\tau M(z) + \zeta} dH(\tau). \end{cases}$$

□

Appendix C: Proof of Theorem 2.3

The convergence of the ESD of ICV has been proved in Theorem 2.2. The rest of Theorem 2.3 is a direct consequence of the following two convergence results.

Lemma C.1. Under the assumptions of Theorem 2.3, we have

$$\lim_{p \rightarrow \infty} 3 \frac{\sum_{i=1}^m |\Delta \bar{\mathbf{Y}}_{2i}|^2}{p} = \zeta, \quad \text{almost surely.}$$

Proposition C.1. Under the assumptions of Theorem 2.3, $F^{\tilde{\Sigma}}$ converges almost surely, and the limit \tilde{F} is determined by \check{H} in that its Stieltjes transform $m_{\tilde{F}}(z)$ satisfies the following equation

$$m_{\tilde{F}}(z) = \int_{\tau \in \mathbb{R}} \frac{1}{\tau (1 - y(1 + zm_{\tilde{F}}(z))) - z} d\check{H}(\tau), \quad \text{for all } z \in \mathbb{C}^+. \quad (\text{C.1})$$

We prove Proposition C.1 first, and then give the proof of Lemma C.1.

Proof of Proposition C.1. We first show the convergence of $F^{\tilde{\Sigma}}$. The main reason that we choose k in such a way that $k/\sqrt{n} \rightarrow \infty$ is to make the noise term negligible. To be more specific, by choosing $k = \lfloor \theta n^\alpha \rfloor$ for some $\alpha \in [(3 + \ell)/(2\ell + 2), 1)$ where ℓ is the integer in Assumption (D.i), we shall show that

$$\tilde{\Sigma} = y_m \sum_{i=1}^m \frac{\Delta \bar{\mathbf{Y}}_{2i} (\Delta \bar{\mathbf{Y}}_{2i})^T}{|\Delta \bar{\mathbf{Y}}_{2i}|^2} \quad \text{and} \quad \tilde{\tilde{\Sigma}} := y_m \sum_{i=1}^m \frac{\Delta \bar{\mathbf{X}}_{2i} (\Delta \bar{\mathbf{X}}_{2i})^T}{|\Delta \bar{\mathbf{X}}_{2i}|^2}$$

have the same LSD. This follow if we can show that

$$\max_{i=1, \dots, m} \left| \frac{|\Delta \bar{\mathbf{Y}}_{2i}|^2}{|\Delta \bar{\mathbf{X}}_{2i}|^2} - 1 \right| \rightarrow 0 \quad \text{almost surely,} \quad (\text{C.2})$$

and

$$y_m \sum_{i=1}^m \frac{\Delta \bar{\mathbf{Y}}_{2i} (\Delta \bar{\mathbf{Y}}_{2i})^T}{|\Delta \bar{\mathbf{X}}_{2i}|^2} \quad \text{and} \quad \tilde{\tilde{\Sigma}} \quad \text{have the same LSD.} \quad (\text{C.3})$$

We start with (C.2). Because

$$\begin{aligned} \left| \frac{|\Delta \bar{\mathbf{Y}}_{2i}|^2}{|\Delta \bar{\mathbf{X}}_{2i}|^2} - 1 \right| &= \left| \frac{|\Delta \bar{\mathbf{X}}_{2i}|^2 + |\Delta \bar{\mathbf{e}}_{2i}|^2 + 2(\Delta \bar{\mathbf{X}}_{2i})^T (\Delta \bar{\mathbf{e}}_{2i})}{|\Delta \bar{\mathbf{X}}_{2i}|^2} - 1 \right| \\ &\leq \left(\frac{|\Delta \bar{\mathbf{e}}_{2i}|}{|\Delta \bar{\mathbf{X}}_{2i}|} \right)^2 + 2 \frac{|\Delta \bar{\mathbf{e}}_{2i}|}{|\Delta \bar{\mathbf{X}}_{2i}|}, \end{aligned}$$

to prove (C.2), it suffices to show

$$\max_{1 \leq i \leq m} \frac{|\Delta \bar{\mathbf{e}}_{2i}|}{|\Delta \bar{\mathbf{X}}_{2i}|} \rightarrow 0 \quad \text{almost surely.}$$

Below, we shall prove the following slightly stronger result:

$$\max_{1 \leq i \leq m, 1 \leq j \leq p} \frac{\sqrt{p} |\Delta \bar{\mathbf{e}}_{2i}^j|}{|\Delta \bar{\mathbf{X}}_{2i}|} \rightarrow 0 \quad \text{almost surely,} \quad (\text{C.4})$$

where for any vector \mathbf{a} , a^j denotes its j th entry.

Note further that for (C.3), using Lemma E.2 in Appendix E, to prove (C.3), it also suffices to show (C.4).

We now prove (C.4). We start with the denominator term $\Delta \bar{\mathbf{X}}_{2i}$. We have $\Delta \bar{\mathbf{X}}_{2i} = \tilde{\mathbf{V}}_i + \tilde{\mathbf{Z}}_i$ for $\tilde{\mathbf{V}}_i$ and $\tilde{\mathbf{Z}}_i$ as defined in (B.3) and (B.4), respectively. Write $\tilde{\mathbf{Z}}_i$ as $\sqrt{w_i} \mathbf{\Lambda} \mathbf{Z}_i$, where w_i is defined in (2.12) and

$$\mathbf{Z}_i = \sum_{|j| < k} \frac{1}{\sqrt{w_i}} \left(1 - \frac{|j|}{k}\right) \int_{((2i-1)k+j-1)/n}^{((2i-1)k+j)/n} \gamma_t d\mathbf{W}_t.$$

By Assumption (D.ii), for all $j \notin \mathcal{I}_p$, Z_i^j are i.i.d. $N(0, 1)$. By using the same trick as in the proof of (3.34) in Zheng and Li (2011), we have

$$\max_{1 \leq i \leq m} \left| \frac{1}{p} |\mathbf{\Lambda} \mathbf{Z}_i|^2 - 1 \right| \rightarrow 0 \quad \text{almost surely.} \quad (\text{C.5})$$

Note that

$$|\Delta \bar{\mathbf{X}}_{2i}|^2 = |\tilde{\mathbf{V}}_i + \tilde{\mathbf{Z}}_i|^2 \geq |\tilde{\mathbf{V}}_i|^2 + |\tilde{\mathbf{Z}}_i|^2 - 2|\tilde{\mathbf{V}}_i||\tilde{\mathbf{Z}}_i|.$$

Assumption (D.iii) implies that for all i , there exists \tilde{C}_1 such that

$$|w_i| \geq \tilde{C}_1 \frac{k}{n}.$$

Therefore, by Assumption (D.vii), there exists $C > 0$ such that

$$|\tilde{\mathbf{Z}}_i|^2 = |w_i| |\mathbf{\Lambda} \mathbf{Z}_i|^2 \geq \frac{C}{p} |\mathbf{\Lambda} \mathbf{Z}_i|^2,$$

which, together with (C.5), implies that there exists $\delta_1 > 0$ such that for all sufficiently large n ,

$$\min_{1 \leq i \leq m} |\tilde{\mathbf{Z}}_i|^2 \geq \delta_1.$$

Moreover, by Assumption (C.ii), $|\tilde{V}_i^j| \leq Ck/n$ for all i, j ; hence, $\max_i |\tilde{\mathbf{V}}_i| = O(\sqrt{p} \times k/n)$, which, by Assumption (D.vii), is $O(\sqrt{1/m}) = o(1)$. Therefore, there exists a constant $\delta > 0$ such that, almost surely, for sufficiently large all n ,

$$\min_{1 \leq i \leq m} |\Delta \bar{\mathbf{X}}_{2i}|^2 \geq \delta. \quad (\text{C.6})$$

It remains to prove that

$$\max_{1 \leq i \leq m, 1 \leq j \leq p} \sqrt{p} |\Delta \bar{\varepsilon}_{2i}^j| \rightarrow 0, \quad \text{almost surely.} \quad (\text{C.7})$$

Observe that if we can show that there exists $C > 0$ such that

$$\max_{1 \leq i \leq m, 1 \leq j \leq p} E|\Delta \bar{\varepsilon}_{2i}^j|^{2\ell} \leq Ck^{-\ell}, \quad (\text{C.8})$$

where ℓ is the integer in Assumption (D.vii), then, for any $\varepsilon > 0$, by Markov's inequality, we have

$$\begin{aligned} P\left(\max_{1 \leq i \leq m, 1 \leq j \leq p} \sqrt{p} |\Delta \bar{\varepsilon}_{2i}^j| \geq \varepsilon\right) &\leq \sum_{1 \leq i \leq m, 1 \leq j \leq p} \frac{p^\ell E|\Delta \bar{\varepsilon}_{2i}^j|^{2\ell}}{\varepsilon^{2\ell}} \\ &\leq \frac{Cmp \cdot p^\ell}{k^\ell \varepsilon^{2\ell}} = O\left(\frac{1}{n^{(2+2\ell)\alpha-2-\ell}}\right), \end{aligned}$$

where the last equation is due to Assumption (D.vii). Because $\alpha > (3+\ell)/(2\ell+2)$ by Assumption (D.vii) again, we have $(2+2\ell)\alpha-2-\ell > 1$; hence, by the Borel-Cantelli Lemma, (C.7) holds.

We now show (C.8), which is a Marcinkiewicz-Zygmund type inequality. We use Theorem 1 in Doukhan and Louhichi (1999) to prove (C.8). For that, we need to verify that $C_{r,2\ell} = O(r^{-\ell})$, where

$$C_{r,2\ell} := \max_{j=1,\dots,p} \max_{1 < M < 2\ell} \sup_{(i_1, \dots, i_{2\ell}) \in \Theta_{r,M,2\ell}} \left| \text{Cov}(\varepsilon_{i_1}^j \cdots \varepsilon_{i_M}^j, \varepsilon_{i_{M+1}}^j \cdots \varepsilon_{i_{2\ell}}^j) \right|,$$

where $\Theta_{r,M,2\ell} = \{(i_1, \dots, i_{2\ell}) : i_1 \leq \dots \leq i_M < i_M + r \leq i_{M+1} \leq \dots \leq i_{2\ell}\}$.

We now verify that $C_{r,2\ell} = O(r^{-\ell})$. For any j and for any $(i_1, \dots, i_{2\ell}) \in \Theta_{r,M,2\ell}$, using the definition of ρ -mixing coefficients we have

$$\begin{aligned} &\left| \text{Cov}(\varepsilon_{i_1}^j \cdots \varepsilon_{i_M}^j, \varepsilon_{i_{M+1}}^j \cdots \varepsilon_{i_{2\ell}}^j) \right| \\ &\leq \rho^j(r) \cdot \sqrt{\text{Var}(\varepsilon_{i_1}^j \cdots \varepsilon_{i_M}^j) \cdot \text{Var}(\varepsilon_{i_{M+1}}^j \cdots \varepsilon_{i_{2\ell}}^j)} \\ &\leq \rho^j(r) \cdot \sqrt{E((\varepsilon_{i_1}^j)^2 \cdots (\varepsilon_{i_M}^j)^2) \cdot E((\varepsilon_{i_{M+1}}^j)^2 \cdots (\varepsilon_{i_{2\ell}}^j)^2)}. \end{aligned}$$

By Hölder's inequality, we have

$$E((\varepsilon_{i_1}^j)^2 \cdots (\varepsilon_{i_M}^j)^2) \leq \left(E(\varepsilon_{i_1}^j)^{2M}\right)^{1/M} \cdots \left(E(\varepsilon_{i_M}^j)^{2M}\right)^{1/M}.$$

And, similarly, for $E((\varepsilon_{i_{M+1}}^j)^2 \cdots (\varepsilon_{i_{2\ell}}^j)^2)$. Because (ε_i^j) s have bounded 4ℓ th moments and $\max_{j=1,\dots,p} \rho^j(r) = O(r^{-\ell})$ by Assumption (D.i), we get $C_{r,2\ell} = O(r^{-\ell})$.

Finally, by using an argument similar to the last part of the proof of Proposition 8 in Zheng and Li (2011) (see pp.3142–3143), we have that $\tilde{\tilde{\Sigma}}$ has the same LSD as

$$\tilde{\mathbf{S}} := \frac{1}{m} \sum_{i=1}^m \mathbf{\Lambda} \mathbf{Z}_i \mathbf{Z}_i^T \mathbf{\Lambda}^T,$$

where \mathbf{Z}_i consists of independent standard normals. It is well known that the LSD of $\tilde{\mathbf{S}}$ is determined by (C.1); hence, by the previous arguments, so is that of $\tilde{\mathbf{\Sigma}}$. \square

We now prove Lemma C.1.

Proof of Lemma C.1. We have

$$\sum_{i=1}^m |\Delta \bar{\mathbf{Y}}_{2i}|^2 = \sum_{i=1}^m |\Delta \bar{\mathbf{X}}_{2i}|^2 + 2 \sum_{i=1}^m \Delta \bar{\mathbf{X}}_{2i}^T \Delta \bar{\boldsymbol{\epsilon}}_{2i} + \sum_{i=1}^m |\Delta \bar{\boldsymbol{\epsilon}}_{2i}|^2.$$

The convergence (C.7) and $p/m \rightarrow y$ imply that $\sum_{i=1}^m |\Delta \bar{\boldsymbol{\epsilon}}_{2i}|^2/p \rightarrow 0$ almost surely. To prove the lemma, it then suffices to show that

$$3 \frac{\sum_{i=1}^m |\Delta \bar{\mathbf{X}}_{2i}|^2}{p} \rightarrow \zeta \quad \text{almost surely,} \quad (\text{C.9})$$

and

$$\frac{\sum_{i=1}^m \Delta \bar{\mathbf{X}}_{2i}^T \Delta \bar{\boldsymbol{\epsilon}}_{2i}}{p} \rightarrow 0 \quad \text{almost surely.} \quad (\text{C.10})$$

We start with (C.9). Write $\Delta \bar{\mathbf{X}}_{2i} = \tilde{\mathbf{V}}_i + \tilde{\mathbf{Z}}_i$ as in the proof of Proposition C.1. The convergence (C.5) implies that

$$\frac{\sum_{i=1}^m |\tilde{\mathbf{Z}}_i|^2}{p} = \sum_{i=1}^m w_i + \text{error},$$

where the error term converges to 0 almost surely. By Riemann integration and Assumption (C.vi) it is easy to show that $\sum_{i=1}^m w_i \rightarrow \zeta/3$, so we get

$$3 \frac{\sum_{i=1}^m |\tilde{\mathbf{Z}}_i|^2}{p} \rightarrow \zeta \quad \text{almost surely.}$$

Furthermore, by using the bound that $\max_i |\tilde{\mathbf{V}}_i| = O(\sqrt{p} \times k/n)$ one can easily show that

$$\frac{2 \sum_{i=1}^m |\tilde{\mathbf{V}}_i| |\tilde{\mathbf{Z}}_i| + \sum_{i=1}^m |\tilde{\mathbf{V}}_i|^2}{p} \rightarrow 0 \quad \text{almost surely.}$$

We therefore get (C.9).

Finally, (C.10) follows from (C.7) and (C.9). \square

Appendix D: Verifying $\widehat{m_{\mathcal{A}_n}}(z) \approx m_{\mathcal{A}_n}(z)$

Suppose that $\widehat{m_{\mathcal{A}_n}}(z) \in D_{\mathcal{A}}(y_n, \sigma_n^2)$
 ($= \{\xi \in \mathbb{C} : z(1 - y_n \sigma_n^2 \xi)^2 - \sigma_n^2(y_n - 1)(1 - y_n \sigma_n^2 \xi) \in \mathbb{C}^+\}$) satisfies the empirical version of equation (2.5); in other words,

$$\widehat{m_{\mathcal{A}_n}}(z) = \int \frac{\tau}{\frac{\tau}{1 - y_n \sigma_n^2 \widehat{m_{\mathcal{A}_n}}(z)} - z(1 - y_n \sigma_n^2 \widehat{m_{\mathcal{A}_n}}(z)) + \sigma_n^2(y_n - 1)} dF^{\mathbf{S}_n}(\tau). \quad (\text{D.1})$$

First, we claim that $\{\widehat{m_{\mathcal{A}_n}}(z)\}$ is tight (regardless of whether $\widehat{m_{\mathcal{A}_n}}(z)$ is the unique solution or not). Suppose to the contrary that $\{\widehat{m_{\mathcal{A}_n}}(z)\}$ is not tight; then, with positive probability, there exists a subsequence $\{n_k\}$ such that $|\widehat{m_{\mathcal{A}_{n_k}}}(z)| \rightarrow \infty$. However, by the tightness of $\{F^{\mathbf{S}_{n_k}}\}$, along such a subsequence, the right-hand side of (D.1) would converge to 0, while the left-hand side blows up, a contradiction.

Next we show that any limit of $\widehat{m_{\mathcal{A}_n}}(z)$ as $n \rightarrow \infty$, denoted by \tilde{m} , has to be $m_{\mathcal{A}}(z)$. Because $\widehat{m_{\mathcal{A}_n}}(z) \in D_{\mathcal{A}}(y_n, \sigma_n^2)$, the limit \tilde{m} satisfies

$$\Im(z(1 - y\sigma^2\tilde{m})^2 - \sigma^2(y - 1)(1 - y\sigma^2\tilde{m})) \geq 0.$$

We want to show that $\tilde{m} \in D_{\mathcal{A}}(y, \sigma^2)$; in other words, the equality sign cannot hold. This can be seen as follows. First, by rewriting (D.1) as

$$\begin{aligned} \widehat{m_{\mathcal{A}_n}}(z) &= (1 - y_n \sigma_n^2 \widehat{m_{\mathcal{A}_n}}(z)) \\ &\cdot \int \frac{dF^{\mathbf{S}_n}(\tau)}{\tau - z(1 - y_n \sigma_n^2 \widehat{m_{\mathcal{A}_n}}(z))^2 + \sigma_n^2(y_n - 1)(1 - y_n \sigma_n^2 \widehat{m_{\mathcal{A}_n}}(z))}, \end{aligned}$$

we see that $(1 - y_n \sigma_n^2 \widehat{m_{\mathcal{A}_n}}(z))$ cannot converge to 0 (because otherwise the right-hand side would converge to 0 while the left-hand side would converge to $1/(y\sigma^2) \neq 0$). It follows by taking the limits on both sides of the equation above that \tilde{m} satisfies

$$\frac{\tilde{m}}{1 - y\sigma^2\tilde{m}} = \int \frac{dF(\tau)}{\tau - z(1 - y\sigma^2\tilde{m})^2 - \sigma^2(y - 1)(1 - y\sigma^2\tilde{m})}.$$

Now, if $\Im(z(1 - y\sigma^2\tilde{m})^2 - \sigma^2(y - 1)(1 - y\sigma^2\tilde{m})) = 0$, then the right-hand side would be a real number; consequently, \tilde{m} has to be a real number as well. However, because we have just proved that $(1 - y\sigma^2\tilde{m}) \neq 0$, if \tilde{m} is a real number, then $\Im(z(1 - y\sigma^2\tilde{m})^2 - \sigma^2(y - 1)(1 - y\sigma^2\tilde{m}))$ cannot be zero, a contradiction.

To sum up, we have shown that \tilde{m} satisfies equation (2.5) and is inside $D_{\mathcal{A}}(y, \sigma^2)$. Therefore, by the uniqueness of the solution to equation (2.5) inside $D_{\mathcal{A}}(y, \sigma^2)$, we have $\tilde{m} = m_{\mathcal{A}}(z)$. Consequently, $\widehat{m_{\mathcal{A}_n}}(z) \approx m_{\mathcal{A}}(z)$ when n (and p) are large.

Finally, because $m_{\mathcal{A}_n}(z) \rightarrow m_{\mathcal{A}}(z)$, we have that $\widehat{m_{\mathcal{A}_n}}(z) \approx m_{\mathcal{A}_n}(z)$.

Appendix E: Two lemmas

Lemma E.1. (Lemma 2.7 in [Bai and Silverstein \(1998\)](#)). Let $\mathbf{X} = (X_1, \dots, X_n)^T$ be a vector where the X_i s are centered i.i.d. random variables with unit variance. Let \mathbf{A} be an $n \times n$ deterministic complex matrix. Then, for any $p \geq 2$,

$$E |\mathbf{X}^T \mathbf{A} \mathbf{X} - \text{tr } \mathbf{A}|^p \leq C_p \left((E|X_1|^4 \text{tr } \mathbf{A} \mathbf{A}^*)^{p/2} + E|X_1|^{2p} \text{tr}(\mathbf{A} \mathbf{A}^*)^{p/2} \right).$$

Lemma E.2. (Lemma 1 in [Zheng and Li \(2011\)](#)). Suppose that for each p , $\mathbf{v}_l = (v_l^1, \dots, v_l^p)^T$ and $\mathbf{w}_l = (w_l^1, \dots, w_l^p)^T$, $l = 1, \dots, m$, are all p -dimensional vectors. Define

$$\tilde{\mathbf{S}}_m = \sum_{l=1}^m (\mathbf{v}_l + \mathbf{w}_l)(\mathbf{v}_l + \mathbf{w}_l)^T \quad \text{and} \quad \mathbf{S}_m = \sum_{l=1}^m \mathbf{w}_l(\mathbf{w}_l)^T.$$

If the following conditions are satisfied,

- (i) $m = m(p)$ with $\lim_{p \rightarrow \infty} p/m = y > 0$,
- (ii) there exists a sequence $\varepsilon_p = o(1/\sqrt{p})$ such that for all p and all l , all the entries of \mathbf{v}_l are bounded by ε_p in absolute value;
- (iii) $\limsup_{p \rightarrow \infty} \text{tr}(\mathbf{S}_m)/p < \infty$ almost surely.

Then, $L(F^{\tilde{\mathbf{S}}_m}, F^{\mathbf{S}_m}) \rightarrow 0$ almost surely, where for any two probability distribution functions F and G , $L(F, G)$ denotes the Levy distance between them.

Appendix F: Two figures in the simulation studies

Figure 7 plots a sample path of (γ_t) .

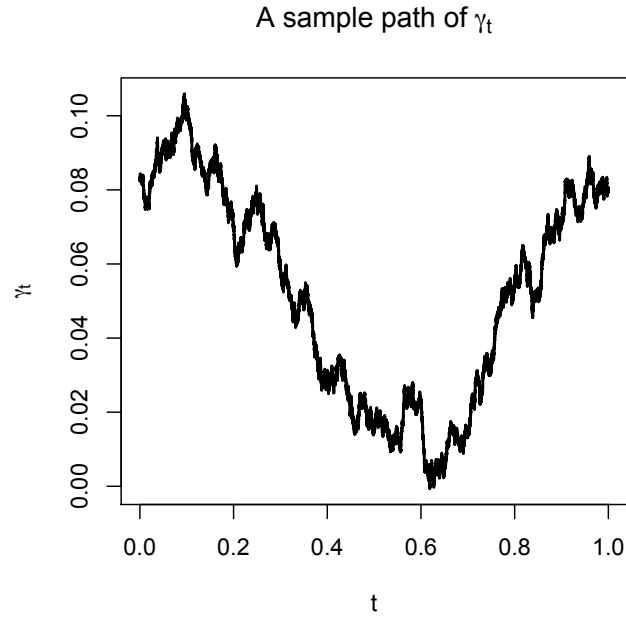
FIG 7. A sample path of the process (γ_t) .

Figure 8 shows a realization of observation times for three stocks under the asynchronous observation setting.

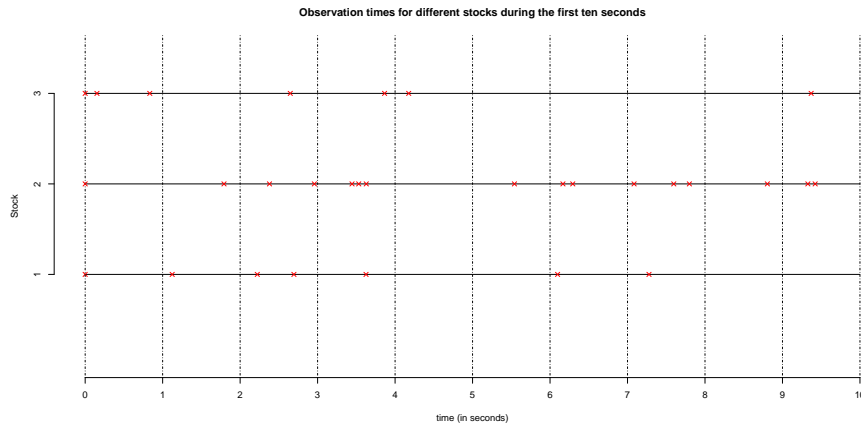


FIG 8. One realization of the observation times during the first 10 seconds for three stocks. The observation times are generated as independent Poisson processes with rate 23,400.

Supplement to “On the inference about the spectral distribution of high-dimensional covariance matrix based on high-frequency noisy observations ”

().

References

- Bai, Z. and Silverstein, J. (1998). “No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices.” *Annals of Probability*, 26, 316–345.
- Bai, Z. and Silverstein, J. (2012). “No eigenvalues outside the support of the limiting spectral distribution of information-plus-noise type matrices.” *Random Matrices: Theory and Applications*, 1, 1150004.
- Doukhan, P. and Louhichi, S. (1999). “A new weak dependence condition and applications to moment inequalities.” *Stochastic Processes and their Applications*, 84, 313–342.
- Dozier, R. and Silverstein, J. (2007b). “Analysis of the limiting spectral distribution of large dimensional information-plus-noise type matrices.” *Journal of Multivariate Analysis*, 98, 1099–1122.
- Silverstein, J. and Bai, Z. (1995). “On the empirical distribution of eigenvalues of a class of large-dimensional random matrices.” *Journal of Multivariate Analysis*, 54, 175–192.
- Xia, N. and Zheng, X. (2017). “On the inference about the spectral distribution of high-dimensional covariance matrix based on high-frequency noisy observations.”
- Zheng, X. and Li, Y. (2011). “On the estimation of integrated covariance matrices of high dimensional diffusion processes.” *Annals of Statistics*, 39, 3121–3151.